GENERAL THEORY OF APPROXIMATION BY FUNCTIONS INVOLVING A GIVEN NUMBER OF ARBITRARY PARAMETERS*

BY

JOHN WESLEY YOUNG

Introduction.

The following is a special case of the problem to be considered in this paper: Given a function $\phi(x)$ of the real variable x, continuous on a finite interval (a, b); to determine the polynomial $p(x)$ of *given* degree n, which gives the closest approximation to the given function ϕ on the interval (a, b). This problem becomes definite only when the meaning of the phrase "closest approximation" has been precisely stated, and the meaning adopted will depend on the ultimate object in view.

TCHEBYCHEV seems to have been the first to consider this problem.† He regarded that polynomial as giving the best approximation, which rendered the maximum of $|p(x) - \phi(x)|$, as x varied over (a, b), as small as possible. A different point of view would lead one to seek a polynomial of the given degree which rendered as small as possible the expression

$$\int_a^b (p - \phi)^2 dx,$$

or the expression

$$\int_a^b |p - \phi| dx,$$

etc. In all of these cases, and in the more general ones to be referred to presently, the problem consists in the determination of a set of parameters a_i of a function $f(x; a_0, a_1, \cdots, a_n)$ of the real variable x, such that the maximum of $|f(x)|$, as x varies over a given interval, shall be as small as possible.

* Presented to the Society April 28, 1906 (under a different title) and September 4, 1906. Received for publication September 4, 1906.

† TCHEBYCHEV, *Sur les questions de minima qui se rattachent à la représentation approchée des fonctions*, Bulletin de la Société phys.-math. de l'Académie impériale des Sciences de St. Pétersbourg, vol. 16 (1858), 145–149; Mémoires de l'Academie impériale de Sciences de St. Pétersbourg, vol. 9 (1859), pp. 201–291.

This problem TCHEBYCHEV himself (loc. cit.) seems to have regarded as the *general type of problem connected with the approximate representation of functions*, although he seems to have done little with this general problem beyond stating it. He confines himself to a detailed discussion of the case where the function f takes the form $p(x)-\phi(x)$, or $r(x)-\phi(x)$, where $p(x)$ is as before a polynomial of given degree, and where $r(x)$ is a rational function of which the degrees of the numerator and denominator are prescribed. The parameters a_i referred to above are then, of course, the coefficients of the various powers of x. The fact that the degrees of the polynomials are prescribed, i. e. the number of arbitrary parameters, is essential to this type of problem.

Certain generalizations of the problems treated by TCHEBYCHEV at once suggest themselves. On the one hand, the functions $p(x)$ or $r(x)$ might be replaced by any function of a given *class* $\mathfrak{C}$ of functions $g(x; a_0, a_1, \cdots, a_n)$. On the other hand, the form of the function f to be considered might be made more general. In the case actually discussed by TCHEBYCHEV, f was simply $g-\phi$, where ϕ was a given continuous function. Both generalizations are included, if we identify f with a function Ug, where Ug denotes the result of operating on g with some *functional operation* U. If then there exists a set of parameters a_i which renders the maximum of $|Ug|$, as x varies over a certain given finite interval, as small as possible, we will call the resulting function g *a function of approximation in the class* $\mathfrak{C}$ *with reference to* U.

The fundamental theoretical problems that now present themselves are as follows: 1) Given an operation U and a class $\mathfrak{C}$, does a function of approximation in $\mathfrak{C}$ with reference to U exist? 2) What are necessary and sufficient conditions that a function in $\mathfrak{C}$ be a function of approximation in $\mathfrak{C}$ with reference to U?

TCHEBYCHEV has given answers to these problems (loc. cit.) for the special cases already attributed to him, where Ug has the form $g-\phi$, and where g is either a polynomial or a rational fractional function of a specified kind. His methods, which lack the degree of rigor required at this day, have recently been revised by KIRCHBERGER,* and still more recently BOREL† has given elegant proofs of those theorems of TCHEBYCHEV which relate to *polynomials* of approximation.

It is the object of the present paper to give a two-fold generalization of TCHEBYCHEV's theorems; the latter will be found as special cases of our theory (§ 5, *a*). The method followed is largely that of BOREL. In § 1 we state the general problem with the necessary precision and prove the existence of a solution under very general conditions. In §§ 2, 3, 4, restricting ourselves to the

* KIRCHBERGER, *Über Tchebychefsche Annäherungsmethoden*, Dissertation, Göttingen, 1902; in abstract, Mathematische Annalen, vol. 57 (1903), pp. 509–540.

† BOREL, *Leçons sur les fonctions de variables réelles, etc.*, 1905, pp. 82–88.

case in which the parameters enter linearly into the functions g and to the case where we have $Ug = g - \phi$, we prove again the existence of a solution, derive necessary and sufficient conditions which a solution must satisfy, and prove that the solution is unique. The results of these sections are summarized in Theorems 4 and 5. Theorem 4 will be found to express a characteristic property of a general class of "functions of approximation," which seems to have been recognized hitherto only for the case of polynomials and rational functions of approximation. After considering in § 5 certain special cases of the theory developed in the preceding sections, we extend the theory further in § 7 to include the case in which Ug has the form $Vg - \phi$, where V is any one-valued, distributive operation. This makes possible the application of the theory to the approximate representation of functions restricted only to satisfy certain functional equations of a general form; some of these are referred to in § 8, where we obtain as an illustration a theorem concerning the polynomial of approximation of given degree for a linear differential equation with constant coefficients. In § 6 we show how the problem may be formulated analytically.

I. The problem in general.

§ 1. *The general existence theorem.*

Let $f = f(x) = f(x;\ a_0, a_1, a_2, \cdots, a_n)$ be a function of the real variable x and $n+1$ parameters a_i, subject to the following conditions A:

$A1$) $f(x;\ a_0, a_1, a_2, \cdots, a_n)$ shall be defined as a one-valued and continuous function of its arguments for every x of a finite interval (a, b) and for all real finite values of the parameters a_i.*

$A2$) For every positive number M there shall exist an N, such that if the relation

$$|f(x;\ a_0, a_1, a_2, \cdots, a_n)| \leqq M$$

be satisfied for all values of x on (a, b), then the parameters a_i all satisfy the relations

$$|a_i| \leqq N \qquad (i = 0, 1, 2, \cdots, n).$$

Now suppose the parameters a_i to be undetermined but fixed. As x varies over (a, b), $|f|$ will attain its maximum value m at least once. This maximum $m(a_0, a_1, a_2, \cdots, a_n)$ by $A1$) is a continuous function of the a_i. Hence as the a_i vary over all real finite values, m certainly has a lower bound μ which is either positive or zero. Our problem is to ascertain whether there exists a

* Throughout this paper the word interval will always imply that the endpoints are included. This condition $A1$) could, moreover, for the purposes of this section be stated more generally by defining $f(x)$ merely on a set of points E on (a, b), with the restriction that E be *perfect*.

finite set of parameters α_i for which this lower bound is actually attained, i. e., such that we have

$$m(\alpha_0, \alpha_1, \alpha_2, \cdots, \alpha_n) = \mu.$$

If such a set exists, we will call it a *minimizing set of parameters for* f. The problem of determining such a set of parameters would seem to arise in connection with most, if not all, problems of approximation by means of functions with a *given* number of arbitrary parameters.

The question concerning the existence of a minimizing set of parameters can be answered in the affirmative. For, if we choose any particular set of parameters, say $a_0', a_1', a_2', \cdots, a_n'$, and denote $m(a_0', a_1', a_2', \cdots, a_n')$ by M, we know that μ is at most equal to M. We can then without loss of generality confine ourselves to functions f such that we have throughout (a, b)

$$|f| \leqq M.$$

But condition $A2$) requires, then, that there exist a number N such that all the parameters of f satisfy the relations

$$|a_i| \leqq N \qquad (i = 0, 1, 2, \cdots, n). \tag{1}$$

Hence the parameters may be restricted to the finite closed domain defined by the relations (1), and in this domain the function $m(a_0, a_1, a_2, \cdots, a_n)$ attains its lower bound μ for at least one set of values $\alpha_0, \alpha_1, \alpha_2, \cdots, \alpha_n$. Hence we have

THEOREM 1. *There exists at least one minimizing set of parameters for any function satisfying conditions* A.

As an immediate corollary of this theorem, we obtain an existence theorem for the very general class of approximation problems referred to in the *Introduction*. Suppose there is given a class of functions, $g(x; b_0, b_1, b_2, \cdots, b_n)$, with the arbitrary parameters b_i, such that a function g is fully determined as soon as the values of the parameters b_i are fixed. Then as we have seen, the problem of determining, on a finite interval, that function g which will give the *best* approximation to a given function (or to a function satisfying a given functional equation) leads to the determination of a minimizing set of parameters for a function Ug, where Ug denotes the function obtained from g by some functional operation U. If then we identify the function Ug with the function f just considered, Theorem 1 gives the desired information concerning the existence of a solution. The parameters b_i of g will in general be the parameters a_i of $f = Ug$. However, it should be noted that in performing the operation U some of the parameters b_i may disappear, and new arbitrary parameters may be introduced. The parameters b_i which actually appear in Ug we will call the *effective* parameters of g with reference to U. Obviously in applying the conditions A to Ug the effective parameters b_i, together with whatever new param-

eters may have been introduced, play the rôle of a_i's. Any function g, the effective parameters of which have the values belonging to a minimizing set for Ug, we call a *function g of approximation with reference to U*. We have, then,

THEOREM 2. *There exists a function g of approximation with reference to the operation U, provided $f = Ug$ satisfies conditions A.*

II. FUNCTIONS OF APPROXIMATION FOR GIVEN FUNCTIONS.

§ 2. *Another form of the existence theorem.*

We now assign to the function f of the preceding section a more explicit but still very general form. We suppose given a set of $n + 1$ functions,

$$s_0, s_1, s_2, \cdots, s_n,$$

of the real variable x. The function s_i we call the elementary function of rank i. By means of some or all of these functions s_j we form a new function,

$$S_k = a_0 s_0 + a_1 s_1 + \cdots + a_k s_k,$$

which we call an S-function of rank k, if s_k is the elementary function of highest rank occurring in S_k. By allowing the parameters a_j to assume all real finite values we obtain the class of S-functions of rank not higher than k, which we will denote by $\mathfrak{S}_k$. It is clear that $\mathfrak{S}_l$ is contained in $\mathfrak{S}_k$, if l is less than k.

We suppose, further, that there is given a function ϕ of x. The function $y = S_n - \phi$ is to play the rôle of the function f of the preceding section. The functions s_j and ϕ are subject to the following conditions B:

$B1$) The functions s_j and ϕ are one-valued and continuous at every point of (a, b).

$B2$) If one parameter a_j is different from zero, the function S_k does not vanish at more than k points of (a, b).

$B3$) For no set of parameters a_j does the function $y = S_n - \phi$ vanish at every point of (a, b).

Condition $B2$) requires a word of explanation. It implies that each point x_1 of (a, b) at which S_k vanishes can be enclosed within a finite interval throughout which (except at x_1) S_k is different from zero. There are then two cases to consider. If S_k changes sign as x passes through x_1, we call x_1 a *simple* zero of S_k; if S_k does not change sign as x passes through x_1, we call x_1 a *double* zero. In applying condition $B2$) *every double zero must be counted twice.**

We restate the discussion in terms of the new symbols as follows. As x varies over (a, b), the function $|y| = |S_n - \phi|$, continuous on (a, b), attains its maximum value m at least once; $m = m(a_0, a_1, a_2, \cdots, a_n)$ is a continuous

* If a or b is a zero of S_k, it counts as simple.

function of the parameters a_j. For we may vary the a_j by a sufficiently small amount to insure that throughout (a, b), where we have

$$y = a_0 s_0 + a_1 s_1 + a_2 s_2 + \cdots + a_n s_n - \phi, \tag{1}$$

y, and hence m, will change by less than ϵ. We seek to determine a function y in $\mathfrak{S}_n$,

$$\sum_n = \alpha_0 s_0 + \alpha_1 s_1 + \alpha_2 s_2 + \cdots + \alpha_n s_n,$$

such that $m(\alpha_0, \alpha_1, \alpha_2, \cdots, \alpha_n)$ shall be equal to the lower bound μ of m; in other words, we seek a minimizing set of parameters for y. If such a minimizing set exists, we call $\sum_n$ a *function of approximation in* $\mathfrak{S}_n$ *for* ϕ.

We have thought it desirable to state these conditions B independently of conditions A, in order that the remainder of the discussion might gain in unity. It is not difficult to show now, however, that conditions B imply A, and hence establish the existence of a solution to our problem. This we shall do by showing that for all functions (1) satisfying a relation $|y| \leqq M$ (M being finite and greater than μ), each of the parameters satisfies a relation

$$|a_j| \leqq N_j,$$

where the N_j are finite numbers.

We note first that a function

$$S_n(x) = \sum_{j=0}^{n} a_j s_j(x)$$

may be determined and is indeed uniquely determined by its values c_h at $n+1$ distinct points ξ_h ($h = 0, 1, \cdots, n$) on (a, b), i. e., by the equations

$$c_h = \sum_{j=0}^{n} a_j s_j(\xi_h) \qquad (h = 0, 1, \cdots, n).$$

For the determinant

$$|s_j(\xi_h)| \qquad (j, h = 0, 1, \cdots, n),$$

does not vanish; otherwise a function S_n, not identically zero, would exist which vanishes at the $n+1$ points ξ_h, in contradiction to $B2$). Denoting by

$$S_n^{(i)}(x) = \sum_{j=0}^{n} a_j^{(i)} s_j(x)$$

that function in $\mathfrak{S}_n$ which has at the ξ_h the values $c_h^{(i)}$ given by the equations

$$c_h^{(i)} = 0 \qquad (i \neq h),$$

$$c_i^{(i)} = 1,$$

we can express any function S_n having any values c_h at ξ_h by the formula*

$$S_n(x) = \sum_{i=0}^{n} c_i S_n^{(i)}(x) = \sum_{i=0}^{n} \sum_{j=0}^{n} c_i a_j^{(i)} s_j(x).$$

* This is a generalization of LAGRANGE's formula of interpolation.

Now any function S_n satisfying the relation

$$|S_n(x) - \phi(x)| \leqq M$$

satisfies the relation

$$|S_n(x)| \leqq M + \lambda,$$

where λ is the maximum of $|\phi(x)|$ on (a, b). But, by the formula just given, every function S_n can be written in the form

$$S_n(x) = \sum_{i=0}^{n} S_n(\xi_i)\, S_n^{(i)}(x),$$

or

$$\sum_{j=0}^{n} a_j s_j(x) = \sum_{j=0}^{n} s_j(x) \sum_{i=0}^{n} S_n(\xi_i)\, a_j^{(i)}.$$

Hence we have

$$a_j = \sum_{i=0}^{n} S_n(\xi_i)\, a_j^{(i)},$$

and so

$$|a_j| \leqq (M + \lambda) \sum_{i=0}^{n} |a_j^{(i)}| \leqq N_j$$

provided

$$|S_n(x) - \phi(x)| \leqq M.$$

This gives the desired condition. Hence we have the following result:

THEOREM 3. *There exists a function of approximation in $\mathfrak{S}_n$ for ϕ, provided conditions B are satisfied.*

§ 3. *A necessary condition for a function of approximation in $\mathfrak{S}_n$.*

We have just shown (Theorem 3) that there exists at least one function of approximation in $\mathfrak{S}_n$ for ϕ under conditions B. Let this function be $\sum_n$, and as before denote $\sum_n$ by y. Then $|y|$ attains its maximum μ in at least one point of (a, b). Let x' be any point for which $y = \mu$ and x'' any point for which $y = -\mu$. The sets $[x']$ and $[x'']$, consisting of all points x' and x'' respectively, are evidently closed since y is continuous on (a, b). Further, we may determine a number 2δ such that the oscillation of y in any interval of length $\leqq 2\delta$ will be less than a positive ϵ, less than μ. The distance of any point x' from the nearest x'' will then be greater than 2δ. Hence *it is possible to divide the interval (a, b) into a finite number of adjoining subintervals,*

$$L_1, L_2, L_3, \cdots, L_p,$$

such that no L_q contains points of both the sets $[x']$ and $[x'']$, and such that one of two consecutive intervals L_q will contain points of $[x']$ and the other will contain points of $[x'']$. Moreover, the points $\zeta_1, \zeta_2, \zeta_3, \cdots, \zeta_{p-1}$ which

divide (a, b) into the intervals L_q must be chosen distinct from any of the points x' or x'', and this evidently is always possible.

The above discussion will apply to any function in $\mathfrak{S}_n$, and it should be noted that the number p of intervals L_q is fixed for a given S_n. But we can now prove the following condition:

A necessary condition that a given function $\sum_n$ in $\mathfrak{S}_n$ be a function of approximation in $\mathfrak{S}_n$ for ϕ is that the number of intervals L_q exceed $n+1$.

We shall show that under the hypothesis $p \leqq n+1$ we can so construct a function S_n in $\mathfrak{S}_n$ that the maximum of $|S_n - \phi|$ on (a, b) is less than μ. Let

$$T_{p-1} = c_0 s_0 + c_1 s_1 + c_2 s_2 + \cdots + c_{p-1} s_{p-1}$$

be a function in $\mathfrak{S}_{p-1}$ which vanishes at each of the points $\zeta_i (i = 1, 2, \cdots, p-1)$. That such a function exists is shown by an argument similar to the one given on p. 336, which makes it clear that only the ratios $c_0 : c_1 : \cdots : c_{p-1}$ are determined by the conditions thus far imposed on T_{p-1}. We may then write

$$T_{p-1} = \eta \cdot T'_{p-1},$$

where T'_{p-1} is fully and uniquely determined, but where η is any arbitrary constant.*

Now, since T_{p-1} vanishes at $p-1$ different points on (a, b) and since by $B2$) it vanishes nowhere else, it follows that T_{p-1}, being continuous, changes sign as x passes from one interval L_q into the next, and only then. We then determine the sign of η so that T_{p-1} is positive within every interval L_q containing a point x' and negative within every interval L_q containing a point x''.

Make every x' the center of an interval I of length δ and every x'' the center of an interval J of length δ, and let K be the set of points on (a, b) not included in any I or J. Then the upper bound μ' of $|\sum_n - \phi|$ on K is certainly less than μ. Now choose $|\eta|$ so small that $|T_{p-1}| = |\eta| |T'_{p-1}|$ is less than the smaller of the numbers $\mu - \epsilon$ and $\mu - \mu'$ at every point of (a, b).

In any interval I we now have

$$\mu - \epsilon < \sum_n - \phi \leqq \mu$$

and

$$0 < T_{p-1} < \mu - \epsilon,$$

since no I contains a point ζ_i. Whence throughout any I the relation

$$0 < \sum_n - T_{p-1} - \phi < \mu$$

is satisfied, and similarly in any interval J

$$0 > \sum_n - T_{p-1} - \phi > -\mu.$$

* If $p = 1$, simply take T_{p-1} equal to η.

Finally, at every point of K we have

$$|\Sigma_n - \phi| \leqq \mu' \qquad \text{and} \qquad |T_{p-1}| < \mu - \mu',$$

and therefore

$$|\Sigma_n - T_{p-1} - \phi| < \mu.$$

Hence the function $S_n = \Sigma_n - T_{p-1}$, which is in $\mathfrak{S}_n$ under the hypothesis $p \leqq n+1$, is such that at every point of (a, b) the relation

$$|S_n - \phi| < \mu$$

holds, which is the desired result.

§ 4. *The sufficiency of the condition and the uniqueness of the solution.*

Conversely, let Σ_n be any function in $\mathfrak{S}_n$ for which the necessary condition of § 3 is satisfied. There exist then at least $n+2$ points

$$x_1 < x_2 < x_3 < \cdots < x_{n+2}$$

on (a, b) at which we have

$$|\Sigma_n(x_\kappa) - \phi(x_\kappa)| = \mu \qquad (\kappa = 1, 2, \cdots, n+2),$$

and

$$\Sigma_n(x_\kappa) - \phi(x_\kappa) = -[\Sigma_n(x_{\kappa+1}) - \phi(x_{\kappa+1})] \qquad (\kappa = 1, 2, \cdots, n+1).$$

Now suppose there were another function

$$\Sigma_n' = \alpha_0' s_0 + \alpha_1' s_1 + \cdots + \alpha_n' s_n$$

in $\mathfrak{S}_n$, such that on (a, b) the maximum value of $|\Sigma_n' - \phi|$ were less than, or even merely equal to, μ. Then the function

$$(\Sigma_n - \phi) - (\Sigma_n' - \phi) = \Sigma_n - \Sigma_n'$$

would be alternately positive (or zero) and negative (or zero) at the points x_κ. But this would require the continuous function $\Sigma_n - \Sigma_n'$ to have on (a, b) at least $n+1$ simple zeros (or their equivalent, if double zeros occur), which by $B2$) is impossible, unless all the parameters of

$$\Sigma_n - \Sigma_n' = (\alpha_0 - \alpha_0') s_0 + (\alpha_1 - \alpha_1') s_1 + \cdots + (\alpha_n - \alpha_n') s_n$$

vanish. But this requires Σ_n and Σ_n' to be identical. Hence we reach, in connection with the result of § 3, the following conclusions:

THEOREM 4. *A necessary and sufficient condition that a function Σ_n in $\mathfrak{S}_n$ be a function of approximation in $\mathfrak{S}_n$ for ϕ is that the number of intervals L_q exceed $n+1$.*

THEOREM 5. *There exists one and only one function of approximation in* $\mathfrak{S}_n$ *for a function* ϕ.*

§ 5. *Special cases of the general theory.*

a) The theorems of Tchebychev on approximation by polynomials and rational functions.

If in the preceding theory we place $s_i = x^i$, that is, if we identify the class $\mathfrak{S}_n$ with the class of polynomials of degree not higher than n, we are led at once to the results of TCHEBYCHEV already referred to in the *Introduction*. Theorems 4 and 5 then give:

There exists one and only one polynomial of approximation of degree n for any function ϕ *on any given finite interval throughout which* ϕ *is continuous; and a necessary and sufficient condition that a given polynomial of degree n be a polynomial of approximation of that degree is that the number of intervals* L_q *exceed* $n+1$.

A similar result was proved originally by TCHEBYCHEV † for the case of a rational function $R_n(x)$, in which the denominator $D(x)$ is fixed and the numerator is of prescribed degree n. We need here only to place $s_i = x^i/D(x)$ to obtain his theorem, which is analogous to the one just given. Conditions B will be satisfied provided no root of $D(x) = 0$ lies on (a, b).

b) Approximation by finite trigonometric series.

If we place $s_i = \cos ix$, our functions S_n assume the form

$$C_n = a_0 + a_1 \cos x + a_2 \cos 2x + \cdots + a_n \cos nx.$$

Such a function satisfies the condition $B2$) concerning the number of zeros, provided the length of the interval (a, b) does not exceed π. For every function C_n can be expressed as a polynomial in $\cos x$ of degree not higher than n. If ϕ also satisfies $B3$), *Theorems* 4 *and* 5 *will apply to the approximate representation of a given continuous function by means of functions of the form* C_n. Other types of trigonometric series having similar properties are as easily obtainable.

c) Other special types of functions having similar properties might be discussed, but it hardly seems worth while to go any further into details. Series of *Legendre polynomials* and series of *hyperbolic functions* are perhaps the simplest that suggest themselves, for which the crucial condition $B2$) can be readily satisfied. The possible application of Theorems 4 and 5 appears to be wide, in view of the general character of the available classes $\mathfrak{S}_n$.

---

*Since the above was written, a special case of this theorem, concerning the representation of functions by finite trigonometric series, has been announced by M. MAURICE FRÉCHET, *Sur l'approximation des fonctions par des suites trigonométriques limitées*, Comptes Rendus de l'Académie des Sciences de Paris, vol. 144, no. 3 (January 21, 1907), p. 124. Cf. also § 5*b*) above.

† Loc. cit.

§ 6. *Analytic formulation of the problem.*

Conditions B are sufficient to give an analytic formulation to the problem of determining the function of approximation in $\mathfrak{S}_n$ for a given continuous function ϕ, in case the functions S_n and ϕ possess derivatives as to x which are one-valued and continuous everywhere except at a finite number of points $\tau_1, \tau_2, \cdots, \tau_\lambda$ on (a, b), with the property that for every τ_i there exists an e_i such that

$$\lim_{x=\tau_i}(x-\tau_i)^{e_i}\cdot\frac{d}{dx}[S_n(x)-\phi(x)]=0.$$

For Theorem 4 implies the existence of $n+2$ points

$$x_1, x_2, \cdots, x_{n+2}$$

on (a, b) for which we have

$$|\Sigma_n(x_i)-\phi(x_i)|=\mu \qquad (i=1, 2, \cdots, n+2).$$

Since at these points $S_n-\phi$ is either a maximum or a minimum, we have also

$$\frac{d}{dx}[\Sigma_n-\phi]=0 \quad \text{when } x=x_i,$$

except possibly when one of the points x_i coincides with $a, \tau_1, \tau_2, \cdots, \tau_\lambda$, or b.

If then we place

$$g(x)=(x-a)(x-\tau_1)^{e_1}(x-\tau_2)^{e_2}\cdots(x-\tau_\lambda)^{e_\lambda}(x-b),$$

a function of approximation in $\mathfrak{S}_n$ would have to satisfy the following system of equations:

$$(S_n-\phi)^2-\mu^2=0, \qquad g(x)\frac{d}{dx}(S_n-\phi)=0, \quad (x=x_i;\ i=1, 2, \cdots, n+2).$$

These $2n+4$ equations contain just $2n+4$ unknowns, namely, the $n+1$ parameters a_j, the $n+2$ quantities x_i, and the quantity μ. Among the sets of parameters satisfying this system must occur the set α_j giving the function of approximation.

III. Functions of approximation for functional equations.

§ 7. *Extension of the theory.*

The results obtained in the previous paragraphs may be extended readily to the problem of finding functions of approximation for a certain general class of functional equations. If we have given a functional equation $U(y)=0$ and a class of functions $g(x; a_0, a_1, \cdots, a_n)$, we define a *function of approximation for* $U(y)=0$ *in the given class* to be a function g which *renders the maximum of* $|U(g)|$ *on* (a, b) *as small as possible.*

Theorems 4 and 5 may be extended to include this type of problem under the following conditions:

We suppose given a set of $m+1$ *elementary* functions,

$$t_0, t_1, t_2, \cdots, t_m,$$

and form, in a way precisely analogous to the one previously used (§ 2), the various classes $\mathfrak{T}_k$ of functions T_k of rank not higher than k,

$$T_k = a_0 t_0 + a_1 t_1 + \cdots + a_k t_k \qquad (k = 0, 1, 2, \cdots, m).$$

We suppose given further a function $\phi(x)$, and a functional operation V which is effective in a single-valued distributive* way on the functions of $\mathfrak{T}_m$, the resulting values $V(T_m)$ being themselves functions of one variable. *Our treatment in* §§ 2, 3, 4 *applies to the case now under consideration in which, in view of the distributive character of* V, *the function* $U(g)$ *has the form*

$$V(T_m) - \phi = a_0 V(t_0) + a_1 V(t_1) + \cdots + a_m V(t_m) - \phi.$$

But this function has the same form as the one already considered in §§ 2, 3, 4, if we identify $V(t_i)$ with s_i, and $V(T_k)$ with S_k. If then we restate conditions B with the substitutions just indicated, and call the new conditions C, we may extend Theorems 4 and 5 to the new type of problem. Before doing so, however, the following remarks should be made regarding the details of the contemplated extension.

If any of the $V(t_i)$ $(i = 0, 1, \cdots, m)$ are zero identically on (a, b), the corresponding parameters a_i are *non-effective* (cf. § 1). Only the *effective* parameters of a T_k play any essential part in the discussion, and hence in stating condition $C2$) (corresponding to $B2$)) the number k must evidently be replaced by $e_k - 1$, where e_k is the number of effective parameters of T_k with respect to V.

If in any function T_k we place equal to zero all the non-effective parameters with respect to V, we call the resulting function *reduced* with respect to V. A reduced function contains just e_k terms and all the parameters occurring in it are effective. In extending our theory we must be careful to specify that the functions T_k with which we operate are to be reduced functions, whenever this is necessary; as, for instance, in the determination of the functions $T_k^{(i)}$ satisfying the relations $V(T_k^{(i)}(\xi_j)) = 1$ or 0, according as we have $i = j$ or $i \neq j$ in the extended existence theorem (cf. p. 336).

* A functional operation A is said to be *distributive*, if we have $A(\alpha+\beta) = A(\alpha) + A(\beta)$ and $A(c\alpha) = cA(\alpha)$, for every constant value c and for every two functions α, β in the realm of definition of A. Cf. PINCHERLE, *Funktionaloperationen und -gleichungen*, Encyklopädie der Mathematischen Wissenschaften, vol. 2 A 11 (1906), p. 767. In our case we have the notion of an operation which is one-valued and distributive *with reference to a class* of functions, nothing being specified *concerning its behavior or existence* for functions not in the class.

The treatment of this more general case, which evidently reduces to the one first considered when V is the identical operation (i. e., $V(T_m) = T_m$), is throughout essentially identical with the treatment already given in detail for the simpler case. The conditions C may now be stated thus:

C 1) The functions $V(t_i)$ $(i = 0, 1, 2, \cdots, m)$ and ϕ are one valued and continuous at every point of (a, b).

C 2) If one effective parameter a_j is different from zero, the function $V(T_k)$ does not vanish at more than $e_k - 1$ points on (a, b).

C 3) For no set of parameters a_j does the function $y = V(T_m) - \phi$ vanish at every point of (a, b).

A function T_m which renders as small as possible the maximum of $|V(T_m) - \phi|$ on (a, b) we call a function of approximation in $\mathfrak{T}_m$ *with reference to* V *and* ϕ. The extended theorems then are as follows:

THEOREM 6. *A necessary and sufficient condition that a function* T_m *in* $\mathfrak{T}_m$ *be a function of approximation in* $\mathfrak{T}_m$ *with reference to* V *and* ϕ *is that the number of intervals* M_q *exceed the number of effective parameters in* T_m.

Here the intervals M_q are constructed with reference to the function $V(T_m) - \phi$, just as the intervals L_q were constructed with reference to the function $S_m - \phi$.

THEOREM 7. *The effective parameters of a function of approximation in* $\mathfrak{T}_m$ *with reference to* V *and* ϕ *are uniquely determined; i. e., there exists one and only one reduced function of approximation in* $\mathfrak{T}_m$ *with reference to* V *and* ϕ.

One or two further details concerning the proof of Theorem 6 require mention. If as in the previous treatment we denote by $\zeta_1, \zeta_2, \cdots, \zeta_{p-1}$ the points which divide (a, b) into the p intervals M_q, it is necessary to construct a function T_l which vanishes at the $p - 1$ points ζ_i and nowhere else on (a, b) (cf. p. 338). This is clearly possible, provided there exists a number $e_l = p$. That such a number does exist is readily seen from the fact that the sequence of numbers e_i $(i = 0, 1, \cdots, m)$ contains each of the integers from 1 to e_m at least once; for we evidently have $e_{i+1} = e_i$ or $e_{i+1} = e_i + 1$ according as a_{i+1} is not or is an effective parameter. The proof, following exactly the lines previously laid down, then leads to the determination of a function T_l satisfying the relation $|V(T_m) - V(T_l) - \phi| < \mu$. But since V is distributive, this implies the relation $|V(T_m - T_l) - \phi| < \mu$, and $T_m - T_l$ under the hypothesis $p \leqq e_m$ is a function in $\mathfrak{T}_m$, which serves to establish the necessity of the condition under consideration. The latter remark applies also to the proof of the sufficiency of the condition. To go through all the details of the proofs of the last theorems appears superfluous.

It is readily seen also that the analytic formulation of the problem given in § 6 may be extended at once to the present case.

8. *Application of the extended theorems.*

In seeking functional equations to which Theorems 6 and 7 apply, we should notice that a particular form of equation is prescribed, viz., $V(y) - \phi = 0$, where y is the unknown function of x and where V must be distributive. Beyond this it is only necessary that conditions C be satisfied. The variety of distributive operations is very large. As examples of interest we may mention those which give rise to linear differential equations with constant or variable coefficients, linear difference equations, linear integral equations, etc.

As a special case, we may consider the following:

Let

$$c_0 \frac{d^\nu y}{dx^\nu} + c_1 \frac{d^{\nu-1} y}{dx^{\nu-1}} + \cdots + c_\nu y - \phi(x) = 0 \tag{1}$$

be a linear differential equation with constant coefficients c_i, $(c_\nu \neq 0)$, and let the class $\mathfrak{T}_m$ be the class of polynomials of degree m. It is clear that conditions C will be satisfied, provided $\phi(x)$ is not identical on (a, b) with any polynomial of degree equal to (or lower than) m. We have then, from Theorems 6 and 7,

THEOREM 8. *There exists one and only one polynomial of approximation of degree m for any linear differential equation* (1) *under the conditions specified, and a necessary and sufficient condition that a given polynomial of degree m be a polynomial of approximation of this degree is that the number of intervals* M_q *exceed* $m + 1$.

PRINCETON UNIVERSITY,
September, 1906.

ON DERIVATIVES OVER ASSEMBLAGES*

BY

E. R. HEDRICK

§ 1. *Introduction.*

The notion of upper and lower derivatives was introduced by DU BOIS-REYMOND and DINI;† they have been extensively studied by these writers, by SCHEEFFER,‡ BAIRE,§ LEBESGUE‖ and others.¶ It is the purpose of this paper to study a concept which will be called the *derivative with respect to an assemblage*, and to extend the known results concerning derivatives, etc.

A proposition which results from this study is that the existence of a *continuous* assemblage derivative insures the existence of the derivative in the usual sense (theorem 4, § 5); this has been made the center of discussion on account of its beauty and applicability. Some further steps are taken, but the evident extensions, some of which are immediate, have been reserved for possible future presentation.

§ 2. *Upper, lower, and assemblage derivatives.*

Given a function $f(x)$ defined** at the points of an interval $a \leqq x \leqq b$,†† we may extend the notions of continuity, oscillation, approach to limits, and so on, by considering only those values of x which belong to any given assemblage E. Consider in particular the ordinary difference quotient

$$Q(x, h) = \frac{f(x+h) - f(x)}{h}.$$

* Presented to the Society (New York) April 28, 1906. Received for publication January 23, 1907.

† See for example DU BOIS-REYMOND, Mathematische Annalen, vol. 16 (1880), p. 115; DINI, German translation: *Grundlagen für eine Theorie der Functionen einer reellen Grösse*, etc.

‡ Acta Mathematica, vol. 5 (1884), pp. 52, 183, 279.

§ BAIRE, *Thèse, Sur les fonctions de variables réelles*, Annali di Mathematica, ser. 3, vol. 3 (1899); and *Leçons sur les fonctions discontinues*, 1905.

‖ LEBESGUE, *Sur l'intégration et la recherche des fonctions primitives*, 1904.

¶ For example, PASCH, Mathematische Annalen, vol. 30 (1887). I shall refer also to BOREL, *Leçons sur les fonctions de variables réelles*, 1905.

** In this paper it will be understood that the functions used are *single valued* at points at which they are said to be defined.

†† The function may be defined, not for all points of an interval, but for points of any assemblage which includes E.

If E is dense at x_0 both to right and left,* four numbers will be obtained by taking the greatest and least limits of $Q(x_0, h)$ as h approaches zero from the right hand or from the left hand respectively; these will be called the upper [or lower] right hand [or left hand] derived numbers,† with respect to E and will be denoted by

$$D^r_{(E)}f(x_0), \qquad D_{r(E)}f(x_0), \qquad D^l_{(E)}f(x_0), \qquad D_{l(E)}f(x_0).$$

For example, if $M[x_0, d, E, Q(x_0, h)]$ denotes the upper limit of $Q(x_0, h)$ for values of h in the interval $0 < h \leqq d$ for which $x_0 + h$ lies in E, then

$$D^r_{(E)}f(x_0) = \underset{d=0}{\text{lower limit}} \{ M[x_0, d, E, Q(x_0, h)] \}$$

If these four derived numbers are all equal, we shall call their common value *the derivative with respect to E*, and we shall denote it by $D_{(E)}f(x_0)$. If E contains no points to the left [right] of x_0, $D^l_{(E)}$ and $D_{l(E)}$ [$D^r_{(E)}$ and $D_{r(E)}$] are meaningless, in which case they are to be neglected in the preceding definition.

We may now state a fundamental theorem:

THEOREM 1. *If $f(x)$ is defined and continuous for all points of an interval $(x_0 - \epsilon, x_0 + \epsilon)$ about $x = x_0$, there exists an assemblage E_r dense at x_0 on the right, for which the derivative of $f(x)$ at x_0 with respect to E_r exists and has any preassigned value λ between $D^r f(x_0)$ and $D_r f(x_0)$, the upper and lower right-hand derived numbers with respect to the continuum, i. e.:*

$$D_{(E_r)}f(x_0) = \lambda, \qquad D^r f(x_0) \geqq \lambda \geqq D_r f(x_0);$$

[or also

$$D_{(E_l)}f(x_0) = \mu, \qquad D^l f(x_0) \geqq \mu \geqq D_l f(x_0).]$$

The truth of this statement is quite obvious from the *continuity* of $Q(x_0, h)$ since $Q(x_0, h)$ actually takes on for some $h < d$ any value between its upper and its lower limits in the interval $0 < h < d$. The cases $\lambda = D_r$ or $\lambda = D^r$ are included even when $D_r = \pm \infty$, etc.

In any case, if $f(x)$ is merely *defined*, an assemblage E_r exists for which $D_{(E_r)}f(x) = D^r f(x)$.‡ And evidently the theorem holds when $D^r f(x)$, etc. are taken with respect to any assemblage E whatever, instead of with respect to the continuum.

If the derivative of $f(x)$ exists at a point $x = x_0$ with respect to each of a *finite* number of assemblages $E_1, E_2, \cdots, E_n$, and has the same value for each

*See DINI, loc. cit., p. 244; BOREL, loc. cit., p. 28; LEBESGUE, loc. cit., p. 67, etc. These definitions are usually given only for the case in which E is the continuum.

†We shall allow the ideal values $\pm \infty$ for D^r, etc., except when specifically excluded. In most instances the word *limit*, without special mention of such possibilities as $D^r = -\infty$, will be used to avoid long circumlocutions.

‡But not necessarily any value *between* $D^r f(x_0)$ and $D_r f(x_0)$.

of them, the derivative of $f(x)$ exists for the total assemblage ΣE formed by combining the given assemblages and has the same value; but this is *not* always true if the number of given assemblages is infinite. In fact, given a continuous function $f(x)$, all the values of x in an interval about $x=x_0$ can always be arranged into a sequence of assemblages $E_1, E_2, \cdots, E_n, \cdots$ for each of which the derivative of $f(x)$ exists and has the same value, which may be chosen at random as any number between $D^r f(x_0)$ and $D_r f(x_0)$ [or D^l and D_l]. *

On the other hand, a necessary and sufficient condition that the derivative of a function $f(x)$ defined in any interval exist with respect to the total assemblage ΣE at a point $x=x_0$, is that the derivative of $f(x)$ exist at $x=x_0$ and have the same value with respect to each of the component assemblages E (finite or infinite in number) and *that the corresponding oscillation* $\omega_{(E)}(x_0, d)$ *approach its limit zero uniformly*, i. e., $|\omega_{(E)}(x_0, d)| < \epsilon$ whenever $|d| < \delta$ where δ is independent of the assemblage E under consideration. Or again a necessary and sufficient condition that $f(x)$ have a derivative at $x=x_0$ with respect to any assemblage H is that the derivative of $f(x)$ should exist and have the same value *for every possible sequence in* H.

§ 3. *The law of the mean.*

Various generalizations of the law of the mean have been stated in terms of the four fundamental derived numbers.† All of these are special cases or corollaries of the following somewhat more general statement:

THEOREM 2. *Let $f(x)$ be defined on an assemblage H. If $f(x)$ is at a maximum with respect to H at a point $x=x_0$ of H, i. e., if $f(x_0) \geqq f(x)$ for all x in H, then*

$$D^l_{(E)} f(x_0) \geqq D_{l(E)} f(x_0) \geqq 0 \geqq D^r_{(E)} f(x_0) \geqq D_{r(E)} f(x_0)$$

where E is any subassemblage of H for which x_0 is a limiting point.

The proof is immediate since $M[x_0, d, E, Q(x_0, h)]$, for example, is not positive for sufficiently small positive values of d; hence the lower limit of M, that is $D^r_{(E)} f(x_0)$, is not positive.

This theorem, combined with theorem 1 and with the well-known theorem that any *continuous* function actually assumes its maximum on any *closed* assemblage, leads to many corollaries. Of these corollaries I shall mention only the following:

* These and the following statements immediately result from theorems on limits for assemblages in general; see for example my paper: *On a Function, etc.*, Annals of Mathematics, 2d series, vol. 7 (1906), p. 177.

† See for example LEBESGUE, loc. cit., p. 70, etc., where a rather complete summary of the usual theorems is given; and BOREL, loc. cit., p. 28.

COROLLARY A. *If $f(x)$ is defined in an interval* $a \leqq x \leqq b$ and if there exists at each point $a \leqq x < b$ an assemblage R_x having points near x to the right and such that $D_{(R_x)}f(x)$ exists and is positive, then $f(x)$ is at a maximum, if at all, at b.*

It is understood that R_x may be chosen differently at each point x. If $f(x)$ is at a maximum at $x = k$, $a \leqq k < b$, then $D_{(R_x)}f(x) \leqq 0$ by theorem 2.

COROLLARY B. (*Rolle's theorem.*) *If $f(x)$ is defined in the interval† $a \leqq x \leqq b$ and if $f(x)$ is at a maximum or at a minimum at $x = x_0$, $(a < x_0 < b)$, then*

$$L \geqq 0 \geqq l$$

where L and l are respectively the upper and lower limits of $D_{(R_x)}f(x)$ [or of $D_{(L_x)}f(x)$] in the interval $a \leqq x \leqq b$, where $R_x[L_x]$ is a random assemblage at the right [left] of x, for every x for which $a \leqq x < b$ $[a < x \leqq b]$.

For if, for example, $D_{(R_x)}f(x) > 0$ (i. e., if $l > 0$), $f(x)$ is at a maximum at $x = b$, by corollary A.

A consideration of the usual function

$$\phi(x) = f(x) - f(a) - \frac{f(b) - f(a)}{b - a}(x - a)$$

leads to an important corresponding generalization of the Law of the Mean, of which a special case is well known:‡

COROLLARY C. (*Law of the Mean.*) *If $f(x)$ is defined and continuous in the interval $a \leqq x \leqq b$, then*

$$L \geqq \frac{f(b) - f(a)}{b - a} \geqq l$$

where L and l are the upper and lower limits, respectively, of the values of $D_{(R_x)}f(x)$ in the interval, where R_x is an assemblage chosen at random as above, and differently for each value of x.

It is now clear that the upper and lower limits of any fixed set of right-hand assemblage derivatives $D_{(R_x)}f(x)$ in any interval are the same as those for any other such set, or for any left-hand set; in particular they are the same

* Instead of interval we may read *perfect assemblage*, in which case the maximum lies at a right-hand boundary point.

† Or in a perfect assemblage, if x_0 is an interior point.

‡ See LEBESGUE, loc. cit., p. 70. The theorem as stated usually applies only to the four fundamental members. The theorem here stated seems not much more general until it is noticed that R_x may be chosen *and then held fixed*, independently at each x. It should be noticed that these theorems really apply to cases in which $D_{(R_x)}f(x)$ does not exist; if it does not we can choose a new R_x, say R'_x, so that $D_{(R'_x)}f(x)$ exists and has any desired value between $D^r_{(R_x)}$ and $D_{r(R_x)}$.

as the upper and lower limits of any one of the four fundamental numbers in the same interval.* For a subinterval (a_1, b_1) surely exists for which $[f(b_1)-f(a_1)]/[b_1-a_1]$ is as near L (or l) for any preassigned set of $D_{(R_x)}f(x)$ as we please. It follows that the upper and lower limits of any fixed set of $D_{(R_x)}f(x)$ as the interval (a, b) enclosing x_0 is diminished toward zero in any manner, that is the maximum and minimum of any fixed set of $D_{(R_x)}f(x)$ at any point $x=x_0$, are always the same; *hence the oscillation of any fixed set of* $D_{(R_x)}f(x)$ *is the same as that of any other fixed set.*†

§ 4. *Determination of a function by given assemblage derivatives.*

A well known method‡ may now be used to prove the following proposition:

THEOREM 3. *If* $D_{(R_x)}f(x)=0$ *for some right-hand assemblage* R_x *at each point* x *of an interval* $a \leqq x \leqq b$ *in which* $f(x)$ *is defined and continuous,* $f(x)$ *is a constant.*

For let

$$\phi(x)=f(x)+\lambda(x-a), \qquad (\lambda>0),$$

$$\psi(x)=f(x)-\lambda(x-a);$$

then $D_{(R_x)}\phi(x)=\lambda>0$ and $D_{(R_x)}\psi(x)=-\lambda<0$ for $a \leqq x < b$; hence $\phi(x)$ is at a maximum at $x=b$ and $\psi(x)$ is at a minimum at $x=b$; that is,

$$\phi(b) \geqq \phi(x) \geqq f(x) \geqq \psi(x) \geqq \psi(b) \qquad (a \leqq x < b),$$

or

$$f(b)+\lambda(b-a) \geqq f(x) \geqq f(b)-\lambda(b-a) \qquad (a \leqq x < b).$$

It follows that $f(x)$ differs from $f(b)$ by at most the arbitrarily small quantity $\lambda(b-a)$.

Since

$$D_{(E)}[f_1(x)+f_2(x)] = D_{(E)}f_1(x) + D_{(E)}f_2(x)$$

if $D_{(E)}f_1(x)$ and $D_{(E)}f_2(x)$ both exist, we may conclude that if $f_1(x)$ and $f_2(x)$ are each continuous in an interval $a \leqq x \leqq b$ and if $D_{(R_x)}f_1(x)=D_{(R_x)}f_2(x)$ where (R_x) is the same for $f_1(x)$ and $f_2(x)$, but is possibly different for any two values of x, then $f_1(x)=f_2(x)+\text{const.}$ In case the ordinary derivative of one of the functions, say $f_2(x)$, exists, we shall have $D_{(R_x)}f_2(x)=df_2(x)/dx$ for any R_x whatever; hence in this case it is sufficient to know that $D_{(R_x)}f_1(x)=df_2(x)/dx$, where R_x is wholly unrestricted.

*Compare DINI, loc. cit., p. 264, etc.; LEBESGUE, loc. cit., p. 71. The usual statements hold only for the four fundamental numbers.

†The obvious consequences of this remark will not be dwelt upon here. It may be remarked that theorem 4 results at once from this; but the total argument necessary is more complex than that which follows. Again the upper and lower integrals of any fixed set of $D_{(R_x)}f(x)$ are evidently equal to those for any other set, in particular for any of the four fundamental numbers and these various functions are integrable whenever any one set of $D_{(R_x)}f(x)$ is so.

‡See DINI, loc. cit., p. 113.

These statements may be somewhat generalized by using a slightly different proof, following that of SCHEEFFER or that of LEBESGUE, who have stated similar theorems *with respect to the four fundamental numbers.** These methods lead to the result: if $f(x)$ is defined and continuous in an interval $a \leqq x \leqq b$ and *if* $D_{(R_x)}f(x) = 0$ *for some assemblage* R_x *which has points near* x *to the right, for all* x *in the interval except at the points of a countable*† *assemblage* E, *then* $f(x)$ *is a constant;* or again, E *need only be of measure zero if we also know that the two right-hand fundamental numbers are not both* $+\infty$ *or both* $-\infty$ *at points of* E.‡ The proofs, being similar to those given by SCHEEFFER and LEBESGUE (l. c.) are here omitted. It should be noticed that *a continuous function is determined by a knowledge of* $D_{(R_x)}f(x)$ *for all except the points of* E. A large part of LEBESGUE's work may be slightly generalized —in some cases the generalization being an essential one—by means of these remarks. One such case will be treated in the next section (p. 352).

§ 5. *Continuous assemblage derivatives.*

From what precedes we may easily deduce the result:

THEOREM 4. *If* $f(x)$ *is defined and continuous in an interval* $a \leqq x < b$, *and if some right-hand assemblage* R_x *exists at every point* x *such that* $D_{(R_x)}f(x) = \phi(x)$ *exists and is continuous, then the ordinary derivative* $df(x)/dx$ *of* $f(x)$ *exists and is equal to* $\phi(x)$; *that is, the existence of a continuous right-hand assemblage derivative insures the existence and continuity of the ordinary derivative.*§

For, if $D_{(R_x)}f(x) = \phi(x)$ exists and is continuous, then the integral

$$\psi(x) = \int_a^x \phi(x)dx,$$

exists, is continuous, and has a derivative in the ordinary sense which is

$$\frac{d\psi(x)}{dx} = \phi(x) \qquad (a < x < b),$$

$$\left.\frac{d\psi(x)}{dx}\right|_{(r)} = \phi(a) \qquad (x = a),$$

where (r) indicates the right-hand derivative in the ordinary sense. Hence $f(x)$ and $\psi(x)$ are two continuous functions for which

*SCHEEFFER, loc. cit., p. 282; DINI, loc. cit., p. 274. Both methods are given by LEBESGUE, loc. cit., pp. 76–78. No essential change in method is necessary.

† What is really proved is that the assemblage E may be any assemblage whose power is less than that of the continuum.

‡ This is slightly more general than to say that not both are infinite.

§ For the similar result for the four fundamental numbers, see, e. g., DINI, loc. cit., p. 267.

$$D_{(R_x)}f(x) = D_{(R_x)}\psi(x) = \frac{d\psi(x)}{dx} \qquad (a \leqq x < \xi < b).$$

It follows that

$$f(x) = \psi(x) + \text{const.} \qquad (a \leqq x \leqq b),$$

and hence $f(x)$ has an ordinary derivative.

Several corollaries suggest themselves. One restatement deserving mention is the following: If a continuous curve $y = \phi(x)$ can be drawn which passes between $y_1 = D^r f(x)$ and $y_2 = D_r f(x)$, i. e., $y_1 \leqq y \leqq y_2$, at every point of an interval $a \leqq x \leqq b$ in which $f(x)$ is continuous, then $f(x)$ has an ordinary derivative equal to $\phi(x)$ in that interval. For theorem 1 shows that this statement is equivalent to theorem 4.

As mentioned in the footnote, p. 349, this theorem also results from the fact that the oscillation of any set of $D_{(R_x)}f(x)$ is the same at each point as the oscillation of any of the four fundamental numbers. Indeed the continuity of any set of $D_{(R_x)}f(x) = \phi(x)$ at a single point $x = x_0$ insures the existence of the ordinary derivative at that point, for the oscillation of $D_{(R_x)}f(x)$ at $x = x_0$ and therefore the oscillation of the function $Q(x_0, h)$ is zero at the point $h = 0$.*

Resuming the argument of § 3 we may write

$$L \geqq \frac{f(\beta) - f(\alpha)}{\beta - \alpha} \geqq l \qquad (a \leqq \alpha < \beta \leqq b),$$

where L and l are the upper and lower limits of $D_{(R_x)}f(x)$ for a continuous function $f(x)$ in the interval $a \leqq x \leqq b$. Therefore

$$\Sigma\,[(\beta - \alpha)L] \geqq \Sigma\,[f(\beta) - f(\alpha)] \geqq \Sigma\,[(\beta - \alpha)l]$$

where the sums indicated are extended over any finite set of subintervals in the interval $a \leqq x \leqq b$. Passing to the limit we have

$$\overline{\int_a^x} D_{(R_x)}f(x)\,dx \geqq f(x) - f(a) \geqq \underline{\int_a^x} D_{(R_x)}f(x)\,dx \qquad (a \leqq x \leqq b),$$

whenever the upper and lower integrals $\overline{\int}$ and $\underline{\int}$ exist, which is surely true, for example, if $D_{(R_x)}f(x)$ is limited, i. e., if $f(x)$ is continuous and not both $D^r f(x)$ and $D_r f(x)$ are $+\infty$ or $-\infty$ at any point. If now

$$\overline{\int_a^x} D_{(R_x)}f(x)\,dx = \underline{\int_a^x} D_{(R_x)}f(x)\,dx,$$

i. e., if $D_{(R_x)}f(x)$ is integrable in the RIEMANN-CAUCHY sense, in which case any assemblage derivative is also integrable and has the same integral, the pre-

* However the derivative may still be discontinuous at $x = x_0$ or even not defined near $x = x_0$ in this case.

ceding inequalities become equations; hence we may say that *if $f(x)$ is continuous in an interval $a \leqq x \leqq b$ and if a right-hand assemblage R_x exists $(a \leqq x < b)$ for which $D_{(R_x)}f(x)$ exists and is integrable, then*

$$\int_a^x D_{(R_x)}f(x)\,dx = f(x) - f(a) \qquad (a \leqq x \leqq b);$$

*that is, the indefinite integrals of $D_{(R_x)}f(x)$ coincide with its primitive functions.**

In order that $\int D_{(R_x)}f(x)\,dx$ exist, $D_{(R_x)}f(x)$ must be continuous except at an assemblage of points of measure zero. Since the ordinary derivative of $f(x)$ exists at every point where $D_{(R_x)}f(x)$ is continuous, it follows that *if $D_{(R_x)}f(x)$ is integrable, $f(x)$ has a derivative except at points of an assemblage E of measure zero, and*

$$\int_a^x D_{(R_x)}f(x)\,dx = \int_a^x \frac{df(x)}{dx}\,dx = f(x) + \text{const.} \qquad (a \leqq x \leqq b),$$

it being understood that the second integral is taken neglecting the points of E. It is easily seen that this is the actual state of affairs whenever $D_{(R_x)}f(x)$ exists and is continuous except at points of an assemblage E whose power is less than that of the continuum; or also whenever $D_{(R_x)}f(x)$ is limited everywhere and is continuous except at points of an assemblage E of measure zero.

Finally let us consider a *fixed* assemblage R at each point x of an interval $a \leqq x \leqq b$ in which $f(x)$ is continuous. For example, let R be a *fixed* sequence, i. e., let R be the sequence

$$x + h_1,\ x + h_2,\ x + h_3,\ \cdots,\ x + h_n,\ \cdots \qquad (h_i > 0),$$

where the sequence $(h_1, h_2, h_3, \cdots, h_n, \cdots)$ is the *same* for every value of x. Then the sequence of continuous functions

$$Q(x, h_1),\ Q(x, h_2),\ \cdots,\ Q(x, h_n),\ \cdots$$

has as its limit $D_{(R)}f(x)$ whenever $D_{(R)}f(x)$ exists; and the upper and lower limits of $Q(x, h_i)$ are in any case equal to the upper and lower limits of $D^r f(x)$ or of any other assemblage derivative, in any interval. Suppose that $D_{(R)}f(x)$ exists for the *fixed* assemblage R; then by a theorem due to BAIRE† $D_{(R)}f(x)$, since it is the limit of a sequence of continuous functions, is pointwise discontinuous‡ on every perfect assemblage. But if $D_{(R)}f(x)$ is continuous, the derivative of $f(x)$ exists in the ordinary sense.

* This generalizes a result given by LEBESGUE, loc. cit., p. 81. See also DUBOIS-REYMOND, loc. cit., p. 115; PASCH, loc. cit., p. 153.

† See BAIRE, *Thèse*, loc. cit., p. 62 or also *Leçons sur les fonctions discontinues*, p. 98. A simpler proof is that given by LEBESGUE; see BOREL, *Leçons*, note II, p. 149.

‡ That is, the function is continuous at least once in every interval.

THEOREM 5. *If there exists any fixed sequence R (or other assemblage) for which $D_{(R)}f(x)$ exists, then that derivative is pointwise discontinuous and moreover the derivative of $f(x)$ exists in the ordinary sense at all points of an assemblage of the second category.* *

For BAIRE has pointed out that a function which is pointwise discontinuous is continuous at points of an assemblage of the second category.

It is not difficult to see that if the above sequence $Q(x, h_i)$ converges to a continuous limit, the convergence is uniform, and the converse is obviously true.† In case it does, the ordinary derivative of $f(x)$ exists, is continuous, and has the same value as $D_{(R)}f(x)$. This convergence is closely allied to what DINI called simple uniform convergence ‡ in the case of any sequence of continuous functions; but the result in this case is much more far-reaching on account of the special nature of sequences which arise from the process of attempted differentiation. For whereas the simple uniform convergence of DINI is sufficient but not necessary for the continuity of the sum of any given sequence, and whereas moreover it is necessary to assume that the given sequence converges, we may announce the result:

THEOREM 6. *A necessary and sufficient condition that the quotient $Q(x, h)$ approach a limit and that that limit be continuous in an interval $a \leqq x \leqq b$ is that $Q(x, h)$ should be simply uniformly convergent § in that interval;* i. e., *that there should exist a fixed sequence $(h_1, h_2, \cdots, h_n, \cdots)$ for which $Q(x, h_i)$ converges uniformly.*

In the preceding statements it is evident that the numbers h_i need not be constant; if they merely are continuous functions of x all the statements made still hold.

COLUMBIA, MO.,
January, 1907.

* An assemblage is of the first category if each point of it belongs to at least one of a *countable* set of non-dense assemblages. Otherwise it is of the second category. BAIRE, *Thèse*, p. 65; *Leçons*, p. 87.

† See my paper in the Annals of Mathematics, 2d series, vol. 7 (1906).

‡ See DINI, loc. cit., p. 147. The statement is equally applicable of course to the convergence of any function $F(x, h)$ as h approaches zero. For if $F(x, h)$ approaches a limit $\phi(x)$ as h approaches zero and if a sequence of values of h can be found for which $F(x, h_i)$ converges uniformly, then $\phi(x)$ is evidently continuous. I shall say that $F(x, h)$ has simple uniform convergence if h_i exist for which $F(x, h_i)$ converges uniformly.

§ This does not here *postulate* that $Q(x, h)$ *itself* approaches any limit whatever.

GEOMETRIE PROIETTIVE DI CONGRUENZA E GEOMETRIE PROIETTIVE FINITE*

DI

BEPPO LEVI

In una nota † presentata a questa Società il 22 Aprile 1905 e pubblicata nelle Transactions nell'Aprile 1906 i sigg. VEBLEN e BUSSEY hanno portata la loro attenzione sopra la possibilità di costruire uno spazio contenente un numero finito di punti ed in cui valga una geometria proiettiva per la quale siano conservate le principali proprietà dell'ordinaria geometria proiettiva lineare. ‡ Essi hanno pure analizzato qualcuno dei caratteri speciali a tali spazi e ne han data una rappresentazione analitica generale, raccogliendo infine, in un ultimo paragrafo, una bibliografia abbastanza vasta di lavori che più o meno hanno attinenza all'argomento trattato.

Pare di qui che agli autori sia sfuggito un cenno assai rapido, ma sufficientemente esplicito, che sopra queste geometrie e sopra altre assai più generali io diedi fin dal 1904 in una Memoria sopra i "*Fondamenti della metrica proiettiva*" §; ond'io chiedo alla Società il permesso di ricordare qui brevemente i punti della mia memoria che si riferiscono alla questione, aggiungendo ancora alcuni sviluppi che non mi paiono privi d'interesse.

La definizione analitica delle geometrie trattate giustificherà pienamente il nome generico che io ho qui attribuito loro di *geometrie proiettive di congruenza.*

---

* Presented to the Society October 27, 1906. Received for publication October 21, 1906.

† *Finite projective geometries.*

‡ Cfr. particolarmente, loc. cit., § 4, pp. 245–247.

§ Memorie della R. Accademia delle Scienze di Torino, ser. 2, vol. 54 (1904); v. il no. 21, *La geometria proiettiva*, pp. 42–44.

Solo dopo che anche il presente articolo era già stato scritto ebbi esatta conoscenza della Nota dello HESSENBERG: "*Ueber die projektive Geometrie,*" Sitzungberichte der Berliner mathematischen Gesellschaft (28 Gennaio 1903), la quale è pur ricordata dai sigg. VEBLEN e BUSSEY (loc. cit., § 9, p. 258). Mi è debito perciò di riconoscere allo HESSENBERG larga parte della priorità qui reclamata, poichè in quella Nota egli afferma con precisione l'indipendenza della geometria projettiva lineare dalla nozione di ordine, e la dimostra per l'appunto colla costruzione di geometrie projettive finite dei tipi che—seguendo i sigg. VEBLEN e BUSSEY—si chiamerebbero (come sarà tosto ricordato) $PG(2, p)$ e $PG(2, p^2)$. Egli rileva pure (come, independentemente, è avvenuto a me) che in geometrie projettive in cui manchi la nozione d'ordine, divengono possibili configurazioni che sono impossibili nell'ordinario spazio projettivo: —così la configurazione d'un quadrangolo completo coi tre punti diagonali allineati, nella $G(2, 2)$ ed anche in piani contenenti infiniti punti (Cfr. il testo, n. II).

Le osservazioni dello HESSENBERG non invadono però per nulla il campo delle ulteriori considerazioni che sono oggetto del presente scritto.

I. Nella citata Memoria, dopo aver stabiliti, in base ai postulati ammessi, i teoremi di Desargues e di Pascal, onde risultò possibile, seguendo il sig. Hilbert, di riferire lo spazio geometrico ad un sistema di coordinate proiettive, io osservava: *

"Mi piace di rilevare come il concetto di ordine degli elementi di una forma di prima specie non abbia avuto fin qui alcuna parte nella istituzione della nostra geometria e non abbia più alcuna ragione di averne in seguito, in tutto lo sviluppo della geometria proiettiva. . . . È essenziale notare che, per questa costruzione della geometria proiettiva, non è necessario che," nella corrispondenza che la rappresentazione per coordinate stabilisce fra i punti degli assi coordinati dello spazio geometrico ed i numeri del campo numerico, "a espressioni aritmeticamente irreduttibili fra loro di questo campo numerico corrispondano punti † diversi dagli assi . . .; bensì è necessario soltanto che a punti diversi corrispondano numeri o espressioni fra loro aritmeticamente irreduttibili. . . ."

Mediante note operazioni geometriche le quali, sugli assi di uno spazio proiettivo, permettono di costruire il punto la cui coordinata relativa all'asse considerato è la somma, la differenza, il prodotto od il quoto delle coordinate relative a due altri punti assegnati ‡ si possono cioè "costruire tutti i punti degli assi corrispondenti agli elementi del campo di razionalità che ha per base un sistema di segni aritmetici (numeri) § corrispondenti a quanti si vogliano punti fissati arbitrariamente sugli assi. Ora si può supporre che due elementi di questo campo, fra loro aritmeticamente irreduttibili, rappresentino tali successioni di operazioni geometriche che conducano allo stesso punto finale. La differenza di questi due elementi aritmetici rappresenterà allora il punto 0" (origine delle coordinate). "L'insieme di tutti gli elementi rappresentanti questo punto 0 costituisce evidentemente un modulo con coefficienti appartenenti al campo di razionalità sopra nominato; e rappresentano uno stesso punto tutte e sole le espressioni di questo campo di razionalità congrue fra loro rispetto a tal modulo. Il prodotto di due punti p, q non sarà mai lo 0 se non è 0 almeno uno di essi." Il modulo nominato sarà cioè primo.

In tutto quanto segue chiamerò questo modulo *il modulo fondamentale della geometria.*

II. I sigg. VEBLEN e BUSSEY si occupano precisamente di questi spazi in cui le coordinate omogenee dei punti sono gli elementi di un campo di Galois $G(p^n)$; sono cioè i sistemi di polinomi in una variabile x a coefficienti appar-

* L. c., p. 42–43.

†Nel testo citato in luogo di punti e assi si parla di raggi e piani coordinati; ciò perchè quelle frasi si riferiscono ad una geometria nella stella.

‡ Cf. HILBERT, *Grundlagen der Geometrie*, § 24–26, ovvero la mia Memoria, no. 17.

§ Forse l'uso che io ho fatto di questa parola "numeri" può aver diminuito la veduta della generalità del concetto nella mente del lettore.

tenenti al campo di integrità naturale,* congruenti rispetto al modulo che ha per elementi fondamentali un numero primo p e un polinomio dato in x di grado n, a coefficienti interi (che si possono sempre supporre $< p$) irreduttibile rispetto al modulo p. Anche per essi i polinomi che costituiscono il modulo rappresentano il punto 0. La geometria di un tale spazio di k dimensioni chiamano una $PG(k, p^n)$.

Si vede come il mio punto di vista comprenda quello dei sigg. VEBLEN e BUSSEY come caso particolare, ma sia assai più generale del loro. E scegliendo convenientemente il modulo fondamentale si può costruire uno spazio proiettivo che, conservando tutte le proprietà lineari che provengono dai soli postulati di appartenenza (*Verknüpfung*), goda ancora di proprietà ben differenti dallo spazio ordinario (in particolare vi manchi la nozione di ordine) *pur possedendo infiniti punti*. Però nemmeno è sfuggito a me il caso particolare notevole presentato dallo spazio con un numero finito di punti:†

"In particolare, se il campo di razionalità considerato è il campo naturale generato dal solo elemento 1, mediante le operazioni di addizione, sottrazione, moltiplicazione e divisione, gli elementi rappresentanti lo 0 saranno tutti quelli della forma $(h/k)n$, dove n è un numero primo e k non è divisibile per n." ‡ Sopra ciascun asse, ciascun punto sarà rappresentato dall'insieme dei numeri razionali che differiscono fra loro per numeri della forma $(h/k)n$, ovvero—e ciò avverrà per un solo punto (il punto all' ∞) su ciascun asse—dall'insieme di tutte le frazioni irreduttibili il cui denominatore è multiplo di n.

Data una frazione irreduttibile p/q il cui denominatore non sia multiplo di n, si possono sempre trovare due numeri interi positivi h, l, tali che $ql - nh = p$ e quindi, $p/q = l - (h/q)n$. Secondo la definizione sopra esposta, l'intero l rappresenterà lo stesso punto che la frazione p/q. Fra i simboli numerici rappresentanti un punto determinato, diverso dal punto all' ∞, di un asse coordinato, esistono dunque sempre dei numeri interi; e poichè rappresenteranno ancora lo stesso punto i numeri l e $l + \lambda n$, ove λ è un numero intero qualunque, positivo o negativo, potranno sempre questi numeri supporsi $\geqq 0$ e $< n$. Ne risulta che la retta possiede nella nostra geometria un numero finito $(n + 1)$ di punti. È il caso che i sigg. VEBLEN e BUSSEY rappresentano con $PG(k, n)$. "Se in particolare si fa $n = 2$, si otterrà uno spazio proiettivo . . ." in cui "il quarto armonico dopo tre elementi dati coincide con uno di questi tre.§

*Il fatto che, alla mia considerazione di un campo di razionalità, si sostituisce qui quella di un campo d'integrità dipende da ciò che i sigg. VEBLEN e BUSSEY si riferiscono ad una rappresentazione per coordinate omogenee, oppostamente a quanto avviene nella mia Memoria.

†Spetta però ai sigg. VEBLEN e BUSSEY il merito di essersi posto il problema di definire *tutti* gli spazi proiettivi finiti e di averlo risolto caratterizandoli nei $PG(k, p^n)$ sopra ricordati.

‡V. la mia *Memoria* citata, p. 43. Nelle linee seguenti della stessa p. 43 si ritrovano le stesse considerazioni che qui seguono, in una esposizione che qui parrebbe meno opportuna al l'intelligenza.

§Cfr. la Nota dei sigg. VEBLEN e BUSSEY, § 3.

L'esempio offerto dal sig. FANO per dimostrare l'indipendenza della proposizione: *il quarto armonico dopo tre elementi dati è distinto da ciascuno di essi*, dai postulati di appartenenza è una realizzazione geometrica di questa costruzione analitica"; onde questo mostra che, in mancanza di quel postulato "molte proposizioni non diverrebbero erronee, ma illusorie."* E la costruzione analitica esposta mostra che tal singolarità non è per nulla legata al *numero* dei punti ammessi alla retta; essa può presentarsi anche in uno spazio d'infiniti punti.†

III. Le geometrie proiettive di congruenza, mentre permettono di scoprire fatti nuovi, conciliabili colle ipotesi fondamentali della geometria proiettiva, danno pur modo di interpretare in un campo reale, anche costituito d'un numero finito di punti, fatti che negli spazi proiettivi ordinarii possono verificarsi solo nel campo complesso.

Come le considerazioni che seguono mostreranno con maggior precisione, questo è, in larga misura, un riflesso naturale della teoria dei numeri algebrici del KRONECKER, per la quale ogni numero algebrico (reale o complesso) viene ad identificarsi con una congruenza fra elementi razionali.

Un notevole esempio si è presentato nella mia *Memoria* nella considerazione di una geometria piana del tipo ultimamente descritto con $n = 3$:‡ il piano proiettivo è allora costituito di 13 punti e si compone di 13 configurazioni di 9 punti identiche alla configurazione dei flessi di una cubica, la qual configurazione, come si sa, è impossibile in un ordinario piano reale.

Sopra un fatto dello stesso ordine hanno portato la loro attenzione i sigg. VEBLEN e BUSSEY: si sa che, sulla retta complessa, cade in difetto la proposizione fondamentale di v. STAUDT che una corrispondenza armonica in cui tre punti coincidano coi loro omologhi è necessariamente l'identità; fu questo il punto di partenza per interessanti ricerche del Prof. SEGRE sopra quelle corrispondenze armoniche che egli chiamò *antiproiettività* e che risultan dal prodotto di trasformazioni proiettive e di una trasformazione per coniugio.§ Ora un fatto analogo si riproduce nelle geometrie finite dei sigg. VEBLEN e BUSSEY. Essi osservano infatti || che i punti d'una retta di una loro geometria si riuniscono in *catene* (analoghe alle catene di v. STAUDT), i punti di ciascuna delle quali si deducono razionalmente da tre punti assegnati qualsiansi della catena medesima: seguendo i nominati autori, se più catene esistono su una stessa

* Cfr. la mia *Memoria* pag. 43–44.

† Cfr. la nota (3) a piè della pag. 43 nella *Memoria* citata.

‡ L. c. no. 31—*Un piano metrico di* 9 *punti e un piano proiettivo di* 13—pp. 53–56: questo esempio presenta una particolare importanza nella dimostrazione della compatibilità e delle dipendenze dei postulati della metrica e di quelli della geometria proiettiva.

§ SEGRE, *Un nuovo campo di ricerche geometriche*, Atti della R. Accademia di Torino, 1890.

|| L. c., § 5, pp. 249, 250.

retta, queste si aggruppano in sistemi più vasti (2-*catene*) contenenti ciascuna tutte e sole le catene, combinazioni lineari di due catene del sistema (e queste due catene hanno, in generale, due punti comuni — distinti o coincidenti); se i punti della retta non appartengono tutti alla stessa 2-catena, esistono ancora sulla retta sistemi lineari più vasti di catene (3-*catene*) e così via: tutti i punti di una retta di una $PG(k, p^n)$ costituiscono in tal modo una *n-catena*. Ora i sigg. VEBLEN e BUSSEY mostrano che al teorema di v. STAUDT si può sostituire quest'altro: *Una corrispondenza armonica sopra una retta di una* $PG(k, p^n)$ *non può avere più di* $n+1$ *punti fissi non appartenenti alla stessa* $(n-1)$*-catena senza ridursi all'identità.**

Questa proposizione stabilisce però per il numero dei punti fissi un limite molto superiore al reale. Io mi propongo di dimostrare, nelle pagine seguenti, che essa può esser sostituita con un'altra assai più precisa.

Mi permetterò anzi di svolgere al riguardo alcune considerazioni generali, da cui si vedrà come si apra qui un largo campo di ricerche algebrico-geometriche.

IV. Sia m una retta di un qualsiasi spazio proiettivo: assegnati arbitrariamente sopra la retta tre punti a, b, c, si può sempre — al più mediante una sostituzione lineare i cui coefficienti siano espressioni razionali nelle coordinate dei tre punti a, b, c — assumere la retta m come asse coordinato e fare in modo che le ascisse dei tre punti a, b, c, siano rispettivamente $0, 1, \frac{1}{0}(=\infty)$.† Se allora ξ è l'ascissa di un punto mobile della retta, $\phi(\xi)$ l'ascissa del punto corrispondente per una determinata trasformazione armonica‡ che tenga fissi i punti a, b, c, è noto§ che la $\phi(\xi)$ deve essere una funzione di ξ che soddisfi alle equazioni funzionali

$$(1) \qquad \begin{cases} \phi(\xi+\eta) = \phi(\xi) + \phi(\eta), \\ \phi(\xi^2) = [\phi(\xi)]^2. \end{cases}$$

Reciprocamente ogni funzione ϕ soddisfacente alle equazioni (1) definisce una trasformazione armonica della retta che muta in sè i punti a, b, c. Invero dalla 1ª delle (1) segue

$$\phi(\xi) = \phi(\xi+0) = \phi(\xi) + \phi(0)$$

onde

$$\phi(0) = 0;$$

$$\phi(\tfrac{1}{0}) = \phi(\xi + \tfrac{1}{0}) = \phi(\xi) + \phi(\tfrac{1}{0})$$

* L. c., § 5, p. 250.

† Dò la preferenza al segno $\frac{1}{0}$ sopra al segno ∞ perchè non nasca alcun dubbio sulla validità dei presenti sviluppi per geometrie finite o, in generale, geometrie nei cui punti non abbiano rappresentazione tutti i numeri reali.

‡ Cioè una trasformazione biunivoca della retta in sè, che muti gruppi armonici in gruppi armonici.

§ DARBOUX, *Sur le théorème fondamental de la géométrie projective*, Mathematische Annalen, vol. 17 (1880), p. 55 (Vedi p. 56-58).

e quindi, poichè non è costantemente $\phi(\xi) = 0$,

$$\phi(\tfrac{1}{0}) = \tfrac{1}{0};$$

infine da

$$\phi(1) = \phi(1^2) = [\phi(1)]^2 \qquad \text{segue} \qquad \phi(1) = 1,$$

e da

$$\phi(0) = \phi(1-1) = \phi(1) + \phi(-1) \qquad \text{segue allora} \qquad \phi(-1) = -1.$$

Si ha inoltre

$$\phi[(\xi+\eta)^2] = \phi(\xi^2 + \eta^2 + 2\xi\eta) = [\phi(\xi)]^2 + [\phi(\eta)]^2 + 2\phi(\xi\eta)$$

$$= [\phi(\xi) + \phi(\eta)]^2 = [\phi(\xi)]^2 + [\phi(\eta)]^2 + 2\phi(\xi)\phi(\eta)$$

onde

$$\phi(\xi\eta) = \phi(\xi)\phi(\eta).$$

Dalle due relazioni

$$\phi(\xi+\eta) = \phi(\xi) + \phi(\eta), \qquad \phi(\xi\eta) = \phi(\xi)\phi(\eta)$$

segue che, se $\xi, \eta, \zeta, \cdots$ sono ascisse di punti qualsivogliano della retta, e $f(\xi, \eta, \zeta \cdots)$ una loro funzione razionale a coefficienti razionali,

$$\phi[f(\xi, \eta, \zeta \cdots)] = f[\phi(\xi), \phi(\eta), \phi(\zeta), \cdots]. \tag{2}$$

In questa relazione (2) sono comprese le (1) come caso particolare: essa può quindi sostituirsi al loro sistema. Se allora ξ, η, ζ, θ sono tali che

$$(\xi\eta\zeta\theta) = -1,$$

la (2) ci dice che

$$\phi(-1) = \phi[(\xi\eta\zeta\theta)] = [\phi(\xi)\phi(\eta)\phi(\zeta)\phi(\theta)],$$

ossia

$$[\phi(\xi)\phi(\eta)\phi(\zeta)\phi(\theta)] = -1,$$

con che si prova che la trasformazione considerata è armonica.

V. Ciò posto si consideri una geometria proiettiva tale che le ascisse dei punti generici della retta m siano funzioni razionali a coefficienti razionali delle ascisse di un certo sistema di punti. Solo per ragion di semplicità e per le applicazioni successive supporremo che questi punti siano in numero finito.* Indicheremo inoltre, d'ora innanzi, i punti della retta m e le loro ascisse cogli

* L'ipotesi del numero finito di punti razionalmente independenti non è per nulla essenziale; tra l'altro potrebbero i coefficienti, anzichè essere razionali, appartenere ad un determinato campo ortoide (KÖNIG, *Einleitung in die allgemeine Theorie der algebraischen Grössen*, Leipzig, Teubner, 1903, p. 7–8), per esempio, esser semplicemente reali. Però l'ipotesi della razionalità dei coefficienti è necessaria per qualcuna delle considerazioni seguenti; qualche altra chiede poche ovvie modificazioni nell'ipotesi contraria. Inoltre molte delle conseguenze che da quanto segue possono trarsi divengono illusorie se non si ammette almeno che l'aggregato dei punti razionalmente independenti possa essere ben ordinato.

stessi simboli. Si potrà, sopra la nostra retta, determinare una base minima

$$0, \frac{1}{0}, 1, x_1, x_2, \cdots, x_r$$

di punti fra loro razionalmente independenti, tali che ogni altro punto sia un'espressione razionale a coefficienti razionali di essi. La relazione (2), insieme colle condizioni $\phi(0)=0$, $\phi(1)=1$, $\phi(\frac{1}{0})=\frac{1}{0}$, che possono considerarvisi incluse, definirà allora la funzione ϕ per tutti i punti della retta tosto ch'essa sia definita nei punti $x_1, x_2, \cdots, x_r$ per modo che, detti $X_1, X_2, \cdots, X_r$ i punti corrispondenti, — il che si potrà rappresentare colla scrittura

$$X_i = \phi(x_i) \qquad (i=1, 2, \cdots, r) \tag{3}$$

— avvenga che:

1°. Ogni relazione algebrica la quale leghi i punti x_i sia pure soddisfatta se alle x_i si sostituiscano le X_i e reciprocamente ogni relazione algebrica fra le X_i sia soddisfatta pure dalle x_i.

2°. I punti

$$0, \frac{1}{0}, 1, X_1, X_2, \cdots, X_r$$

costituiscano a lor volta una base razionale (che sarà necessariamente minima) pei punti della retta, per modo che ciascuno di questi si esprima razionalmente mediante essi.

Nella condizione 1° è evidentemente inclusa quest'altra, che:

3°. Se un medesimo punto si esprime razionalmente in due modi differenti mediante le x_i (o le X_i) le due espressioni che risultano dopo la sostituzione (3) (o la sua inversa) risultino ancora equivalenti.

Invero, l'equivalenza di due espressioni razionali $f_1(1, x_1, x_2, \cdots, x_r)$, $f_2(1, x_1, x_2, \cdots, x_r)$ si traduce in una relazione algebrica, $f_1 - f_2 = 0$, fra le x_i: allora, a causa della condizione 1°, sarà pure

$$f_1(1, X_1, X_2, \cdots, X_r) - f_2(1, X_1, X_2, \cdots, X_r) = 0.$$

VI. Il supporre che fra gli elementi di una base razionale minima passi una relazione algebrica $F(1, x_1, x_2, \cdots, x_r)=0$, equivale a considerare una geometria di congruenza; equivale cioè a dire che l'espressione $F(1, x_1, x_2, \cdots, x_r)$ rappresenta lo 0 ed appartiene quindi al modulo fondamentale della geometria.* Inversamente una geometria di congruenza il cui modulo fondamentale sia interamente costituito soltanto da polinomi si può considerare come rappresentante di una geometria avente una base razionale, fra i cui elementi sussistano le relazioni algebriche che si ottengono uguagliando a 0 tali polinomi. La cosa è differente quando del modulo fondamentale faccia parte un numero (necessaria-

* È questo il concetto fondamentale della teoria di KRONECKER dei numeri algebrici.

mente intero e primo); e tal differenza appare ben naturale se appena si osserva che, nel primo caso, l'ammissione di un modulo fondamentale equivale ad ammettere che taluno dei punti razionalmente independenti è però scelto per modo che da questi si può ottenere il punto 0 mediante una determinata successione di operazioni razionali (proiezioni e sezioni), mentre nel secondo caso l'ammissione del modulo fondamentale influisce effettivamente sul concetto di retta (e cioè sul concetto delle operazioni razionali fondamentali: proiezioni e sezioni), per modo che per una determinata successione di tali operazioni, effettuate su punti fissati *a priori*, perchè le dette operazioni abbiano senso determinato (i punti $0, \frac{1}{0}, 1$), riconduce da essi al punto 0.* In ogni caso, basta ritornare sul significato e sulla deduzione della (2) per riconoscere che, nel caso delle geometrie di congruenza, in essa l'uguaglianza va ora sostituita con una congruenza rispetto al modulo fondamentale. Si chiami M questo modulo, e si osservi che ciascun punto X_j sarà rappresentato da una classe di espressioni razionali nelle x_i, congrue fra loro rispetto al modulo M; il problema della determinazione di una trasformazione ϕ equivale allora a *determinare una sostituzione*

$$X_j \equiv \psi_j(x_i) \pmod{M} \qquad (i, j = 1, 2, \cdots, r) \tag{4}$$

tale che

1° *trasformi ogni polinomio del modulo M (in cui alle variabili x_i si pensino sostituite le corrispondenti X_i) in un polinomio del modulo medesimo.*

2° *i punti $0, \frac{1}{0}, 1, X_1, \cdots, X_r$ costituiscano una base razionale dei punti della retta.*

La prima condizione mostra che il modulo fondamentale della geometria, rispetto alla base $0, \frac{1}{0}, 1, X_1, \cdots, X_r$, si otterrà sostituendo semplicemente le lettere X alle x nei polinomi del modulo fondamentale relativo alla base $0, \frac{1}{0}, 1, x_1, \cdots, x_r$; indicando con $\overline{M}$ il risultato di questa sostituzione, la condizione 2°si potrà enunciare chiedendo che *la trasformazione* (4) *abbia una inversa univocamente determinata*

$$x_i \equiv \Psi_i(X_j) \pmod{\overline{M}}. \tag{4'}$$

VII. Le x_i e le X_j erano fin qui simboli rappresentanti punti razionalmente independenti della nostra retta. Si interpretino ora per un istante come variabili. Quando del modulo M non faccia parte un numero, la condizione ch'esso sia primo (n°. I) si traduce nel fatto che appartengano ad esso tutti e soli i polinomii che, uguagliati a 0, determinano, nello spazio di coordinate x_i, iper-

* I casi già ricordati della $G(2, 2)$ in cui esiste una *terna* armonica di punti e della $G(2, 3)$ in cui esiste la configurazione dei 9 flessi d' una cubica, chiariscono in modo evidente questa osservazione. Come in un piano di soli punti razionali (e quindi reali), i punti delle nostre geometrie si ottengono con sole operazioni razionali (projezioni e sezioni) dai punti di coordinate $0, \frac{1}{0}, 1$: ma punti che per tal generazione risulterebbero necessariamente distinti nell' ordinario piano reale, vengono ora a coincidere.

superficie passanti per una determinata varietà irreduttibile (razionalmente).* Segue allora che *la trasformazione* (4), *in cui si sostituisca il segno* $= al \equiv$, *rappresenterà una trasformazione algebrica di questa varietà in sè*, la quale, a causa della esistenza di una inversa univoca (4′) *sarà pure birazionale.* Inversamente ogni trasformazione birazionale a coefficienti razionali della varietà rappresentativa del modulo fondamentale in sè definirà una sostituzione (4) la quale muterà una ipersuperficie per essa varietà in una analoga ipersuperficie, e quindi, operata sopra un polinomio qualunque del modulo, lo muterà in un polinomio del modulo. E questa sostituzione ammetterà una inversa (4′): per ottenerla si considerino le espressioni razionali delle x nelle X che, *sulla varietà rappresentativa del modulo* M, definiscono la trasformazione inversa della (4): saranno esse i secondi membri Ψ della sostituzione (4′): invero il prodotto delle due trasformazioni $X_i = \psi_i(x_j), x_j = \Psi_j(X_i')$ sarà una trasformazione $X_i = \Theta_i(X_k')$ la quale, quando le X_k' si interpretano come coordinate di un punto generico della varietà rappresentativa del modulo, deve ridursi all'identità, cosicchè deve essere

$$\Theta_i(X_k') \equiv X_k' \pmod{M'}$$

dove M' è il modulo M scritto nelle variabli X' in luogo delle x.† Quindi *ad ogni trasformazione birazionale a coefficienti razionali della nominata varietà in sè corrisponde una trasformazione armonica della nostra geometria.* Varrà quindi il teorema di v. STAUDT quando la varietà rappresentativa del modulo fondamentale non ammetta trasformazioni birazionali a coefficienti razionali in sè.

* Invero, non potrebbero tutte queste ipersuperficie avere a comune una varietà riduttibile, altrimenti apparterrebbero al modulo polinomi prodotti di altri non appartenenti ad esso, in quanto rappresentano ipersuperficie passanti per le singole parti soltanto di essa varietà. D'altra parte, per un noto teorema del sig. HILBERT (*Ueber die Theorie der algebraischen Formen*, Mathematische Annalen, vol. 36 (1890), p. 474; KÖNIG, *Einleitung in die allgemeine Theorie der algebraischen Grössen*, p. 366) tutti i primi membri delle equazioni delle ipersuperficie per una varietà determinata costituiscono un modulo. Nè, nel caso nostro, potrebbero le forme del modulo M corrispondere a una parte soltanto delle ipersuperficie per la detta varietà comune, e nemmeno potrebbe tal varietà comune venir a mancare, poichè ancora per un teorema dello HILBERT (*Ueber die vollen Invariantensysteme*, Mathematische Annalen, vol. 42 (1893), p. 320; KÖNIG, l. c, p. 399) ed uno del LASKER (*Zur Theorie der Moduln und Ideale*, Mathematische Annalen, vol. 60 (1905), p. 30; SEVERI, *Su alcune proprietà dei moduli di forme algebriche*, Atti della R. Acc. delle Sc. di Torino, 1906), una conveniente potenza di un polinomio qualunque nel secondo caso, o rappresentante un' ipersuperficie per la varietà considerata nel primo, deve appartenere al modulo: ma il modulo non sarebbe primo (o, più direttamente, non sarebbe soddisfatta la proprietà fondamentale dei nostro campo numerico, che un prodotto sia $= 0$ solo se è nullo qualcuno dei fattori) se al modulo potesse appartenere una potenza di un polinomio, e non il polinomio medesimo.

† Invero, posto $\Theta_i(X_k') = \theta_i(X_k')/\theta_i'(X_k')$, il polinomio $\theta_i(X_k') - X_i'\theta_i'(X_k')$ dovrà rappresentare una ipersuperficie passante per la varietà considerata e dovrà quindi appartenere al modulo M'; quindi $\theta_i(X_k') \equiv X_i'\theta_i'(X_k') \pmod{M'}$ e, poichè certamente $\theta_i'(X_k')$ non appartiene ad M', $\Theta_i(X_k') \equiv X_i' \pmod{M'}$.

Una interpretazione delle condizioni 1° e 2° nel caso che il modulo contenga un numero presenta assai maggiori difficoltà.

VIII. Ci volgiamo ora al problema da cui siamo partiti: *Supponendo di considerare una geometria finita, in cui quindi il modulo fondamentale sia della forma*

$$M = [p, F_n(x)]$$

ove p è un numero primo e $F_n(x)$ un polinomio in x a coefficienti interi, di grado n e irreduttibile rispetto al modulo p, qual è il numero massimo di punti $\xi_1, \xi_2, \cdots, \xi_\nu$ tali che il verificarsi della relazione $\phi(\xi_k) = \xi_k$ per ogni $k \leq i$ non abbia per conseguenza $\phi(\xi_{i+1}) = \xi_{i+1}$, ma il verificarsi dell'uguaglianza*

$$\phi(\xi_k) = \xi_k$$

per tutti questi punti abbia per conseguenza che la trasformazione armonica definita dalla ϕ è l'identità?

In tal modo deve infatti interpretarsi la generalizzazione del teorema di v. STAUDT, perchè è ben facile vedere che esistono sulla retta punti tali che ogni trasformazione armonica ϕ che tenga fermi al solito i punti $0, \frac{1}{0}, 1$ e uno di essi è senz'altro l'identità: tale è per esempio il punto x, dove x è la lettera ordinatrice dei polinomi, o, se si vuole, una radice primitiva del campo di Galois; e solo può dubitarsi che possano esistere trasformazioni armoniche ϕ le quali spostino questi punti, tenendo però fermi altri punti della retta, oltre i punti della catena che contiene i punti $0, \frac{1}{0}, 1$. Per risolvere tal questione occorre che ricordiamo anzitutto alcune proprietà dei campi di Galois:

Se ξ_1 è un elemento di un campo di Galois $G(p^n)$ che non sia un numero intero, l'insieme dei polinomi interi in ξ_1, ridotti rispetto al modulo M che definisce il campo, costituisce un campo di Galois $G(p^{n'})$ [che si potrà indicare con $(1, \xi_1)$] contenuto in $G(p^n)$, ed il numero n' è divisore di n. Se poi ξ_2 è un nuovo elemento di $G(p^n)$ non appartenente a $G(p^{n'})$, l'insieme dei polinomi interi in ξ_1, ξ_2, ridotti rispetto al modulo M, costituirà un nuovo campo di Galois $G(p^{n''})$ [che si potrà indicare con $(1, \xi_1, \xi_2)$] contenente $G(p^{n'})$ e contenuto in $G(p^n)$ per modo che n'' sarà divisore di n e n' divisore di n''.†

Si osservi ora che se una trasformazione armonica ϕ trasforma ξ_1 in se

*Sempre quando al modulo appartiene un numero p (necessariamente primo) tutti i coefficienti dei polinomi che costituiscono gli elementi del campo numerico considerato possono supporsi numeri interi per le osservazioni già fatte al No. II.

†Cfr. DICKSON, *Linear groups with an exposition of the Galois Field theory*, Leipzig, 1901, pp. 50 e 51. Per riconoscere l'esattezza di queste affermazioni basta osservare che se λ è un elemento non appartenente al campo di Galois $G(p^\nu)$, un campo di Galois contenente λ e $G(p^\nu)$ conterrà almeno i $p^{\nu+1}$ elementi somme degli elementi di $G(p^\nu)$ cogli elementi $h\lambda$ $(0 \leqq h < p)$, e se, oltre queste somme, contiene un altro elemento μ, ne conterrà pure $p^{\nu+2}$ ottenuti combinando in simil modo μ con questi $p^{\nu+1}$ e così via.

stesso, trasformerà in sè tutti gli elementi del campo $(1, \xi_1)$, e se trasforma in sè ancora ξ_2, trasformerà pure in sè tutti gli elementi del campo $(1, \xi_1, \xi_2)$ e così via. Ne risulta che *il numero dei punti ξ_k definiti nel precedente enunciato è minore o uguale al numero dei termini della massima successione di numeri*

$$n', n'', n''', \cdots, n$$

$(>1 \text{ e } \leqq n)$, *ciascuno dei quali sia divisore del seguente;* altrimenti detto, *è minore o uguale al numero dei divisori primi di n.*

Ma si può aggiungere che *questo massimo è sempre raggiunto*, cosicchè al "minore od uguale," si potrà sostituire "*uguale*" senz'altro. Si osservi infatti che, nel caso presente in cui gli elementi della nostra geometria analitica appartengono ad un campo di Galois $G(p^n)$, le condizioni 1° e 2° imposte alla trasformazione (4) equivalgono a dire che

$$X \equiv x^{p^\nu} \pmod{M} \qquad (0 \leqq \nu < n). \tag{5}$$

Invero la X, dovendo essere radice della congruenza $F_n(x) \equiv 0 \pmod{M}$, dovrà avere la forma (5)* e d'altronde l'inversa della (5) sarà data evidentemente da $X^{p^{n-\nu}} \equiv x^{p^n} \equiv x \pmod{M}$.† Si corrisponderanno per la trasformazione ϕ gli elementi $x^{\mu p^\nu}$ e X^μ; e sarà

$$X^\mu \equiv x^{\mu p^\nu} \equiv x^\mu$$

quando

$$\mu(p^\nu - 1) \equiv 0 \pmod{p^n - 1}.$$

Ora il massimo comun divisore di $p^\nu - 1$ e $p^n - 1$ è della forma $p^l - 1$ ove $\nu = ll'$, $n = l\lambda$ ed l' e λ sono primi fra loro; ed il più piccolo valore di μ per cui questa congruenza è soddisfatta rende $\mu(p^l - 1) = p^n - 1$. Reciprocamente, scelto arbitrariamente l divisore di n e posto $n = l\lambda$, l' primo con λ, $\nu = ll'$, esiste un $\mu = (p^n - 1)/(p^l - 1)$ e quindi $< p^n - 1$ tale che $\mu(p^\nu - 1) \equiv 0 \pmod{p^n - 1}$ e quindi $X^\mu \equiv x^\mu$. Segue che condizione necessaria e sufficiente affinchè la trasformazione (5) lasci fermo qualche punto ad ascissa non numerica è che ν ed n abbiano divisori comuni: se l è il massimo comun divisore di ν ed n resteranno fissi tutti i punti del campo di Galois $(1, x^\mu)$ ove $\mu = (p^n - 1)/(p^l - 1)$: sarà questo un $G(p^l)$.

Fissata quindi arbitrariamente una qualunque successione

$$n', n'', n''', \cdots, n \qquad (n' > 1)$$

di divisori di n tali che ciascuno di essi sia uguale al precedente moltiplicato per un numero primo, si può ad essa far corrispondere una successione di campi di Galois

$$G(p^{n'}), G(p^{n''}), G(p^{n'''}) \cdots, G(p^n)$$

---

*SERRET, *Cours d'algèbre supérieure*, 5me édition, 1885, T. 2, p. 180.

†Cf. DICKSON, l. c., p. 11; SERRET, l. c., No. 346, p. 132.

dei quali l'ultimo sia il campo totale dei punti della nostra retta, e ciascuno dei quali contenga tutti i precedenti; e a ciascuno $G(p^{n^{(k)}})$ *di questi campi si possono far corrispondere trasformazioni* $\phi^{(k)}$ *tali che tengano fermi tutti i punti di* $G(p^{n^{(k)}})$ *(e quindi i punti dei campi precedenti) ma spostino ogni punto dei campi successivi, che non appartenga a questi. Una data* ϕ *sarà una* $\phi^{(k)}$ *tosto che essa tenga fermo un punto di* $G(p^{n^{(k)}})$*, non appartenente a* $G(p^{n^{(k-1)}})$*, e sposti un punto di* $G(p^{n^{(k+1)}})$*. Dato* $n^{(k)}$ *il numero delle* $\phi^{(k)}$ *differenti è il numero dei numeri primi con* $n/n^{(k)}$ *e minori di esso.*

Per assicurare che una ϕ *si riduce all'identità occorre verificare al più che essa tien fissi tanti punti, ciascuno razionalmente independente dai precedentemente considerati, quanti sono i divisori primi del numero* n*.*

Particolarmente notevole è il caso in cui n sia primo; allora, qualunque sia l'elemento ξ di $G(p^n)$, non numero, il campo $(1, \xi)$ è il campo totale $G(p^n)$; quindi se $\phi(\xi) = \xi$, sarà, per ogni altro elemento η, $\phi(\eta) = \eta$. *Se* n *è primo una trasformazione armonica che, oltre ai tre punti* $0, 1/0, 1$ *tenga pur fermo un punto qualunque non appartenente alla catena di questi tre sarà sempre l'identità.*

TORINO, ITALIA, 13 *Ottobre* 1906.

COLLINEATIONS IN A FINITE PROJECTIVE GEOMETRY*

BY

OSWALD VEBLEN

In the paper by Mr. BUSSEY and myself on *Finite Projective Geometries* it was shown † that in a $PG(k, p^n)$ any transformation of the form

$$\frac{x_i'}{x_{k+1}'} = \frac{a_{i1}x_1^{p^m} + a_{i2}x_2^{p^m} + \cdots + a_{ik+1}x_{k+1}^{p^m}}{a_{k+11}x_1^{p^m} + a_{k+12}x_2^{p^m} + \cdots + a_{k+1k+1}x_{k+1}^{p^m}} \qquad (i=1, \cdots, k), \tag{1}$$

where m is zero or an integer less than n, is a collineation. As a result of Dr. LEVI's article ‡ it is possible to prove the converse proposition, namely, that *every collineation in* $PG(k, p^n)$ *is of type* (1). The following argument connects this theorem directly with our former article. It will be sufficient to give the argument for the case, $k = 2$.

A projective collineation, § or, in other words, a linear transformation, is determined by the quadrangle into which it transforms the quadrangle (0 0 1), (1 1 1), (0 1 1), (1 0 1). It follows that an arbitrary collineation is the product of a projective collineation by a collineation which leaves these four points invariant. The latter type of collineation may be called *antiprojective*, in the language of SEGRE. ‖ Its existence is proved by the existence of transformations of type (1) where $m \neq 0$.

An antiprojectivity, by its definition, leaves invariant the points of the x_1 axis for which x_1/x_3 is 0, 1, or ∞, and, therefore, all points of the chain determined by these three, namely, all points whose coördinates are integers modulo p. From the quadrangle-construction for addition and multiplication and the fact that a collineation transforms a complete quadrangle into a complete quadrangle it follows that if three points of the x_1 axis for which x_1/x_3 is a, b, and c respectively are so related that

* Presented to the Society December 28, 1906. Received for publication November 16, 1906.

† These Transactions, vol. 7 (1906), p. 252.

‡ Page 354 of this number of the Transactions. To our former list of works relating to finite geometries we add references to this paper of LEVI, *Geometrie proiettive di congruenza e geometrie proiettive finite*, and to an earlier paper, *Fondamenti della metrica proiettiva*, Memorie della R. Accademia delle Scienze di Torino, ser. 2, vol. 54 (1904), and also to a book by G. ARNOUX, *Arithmétique Graphique. Les espaces arithmétiques hypermagiques*, Paris, 1894. ARNOUX applied his "arithmetic spaces," which may be identified with our $EG(k, p)$, to construction of magic squares.

§ These Transactions, vol. 7 (1906), p. 253.

‖ Cf. LEVI's remarks on page 357.

$$a+b=c \qquad \text{or} \qquad ab=c,$$

then the points, with coördinates $\phi(a)$, $\phi(b)$, $\phi(c)$, into which they are transformed are so related that

$$\phi(a)+\phi(b)=\phi(c) \qquad \text{or} \qquad \phi(a)\phi(b)=\phi(c).$$

In other words the transformation of the points of the x_1 axis effected by an antiprojectivity is subject to the conditions:*

$$\phi(a)+\phi(b)=\phi(a+b); \tag{2}$$

$$\phi(a)\cdot\phi(b)=\phi(ab); \tag{3}$$

$$\phi(0)=0, \qquad \phi(1)=1, \qquad \phi(\infty)=\infty. \tag{4}$$

If e is a primitive root of the $GF[p^n]$ then every mark of the field may be written in the form

$$e^k.$$

But by (3)

$$\phi(e^k)=[\phi(e)]^k.$$

Hence if $\phi(e)=e^m$, then $\phi(e^k)=e^{mk}=(e^k)^m$ and

$$\phi(x)=x^m.$$

If k is a primitive root of the $GF[p]$ then

$$\phi(k)=k.$$

Hence m must be of the form p^l† and therefore, as a transformation of the points of the x_1 axis, an anti-projectivity is given by the equation

$$\frac{x_1'}{x_3'}=\frac{x_1^{p^l}}{x_3^{p^l}},$$

and the number of antiprojectivities, exclusive of the identity, is $n-1$.

An antiprojectivity is, however, completely determined by the permutation it effects among the points of the x_1-axis. For by projection from the center (1 0 1) this permutation determines the only possible permutation of points on the fixed line $x_1=x_2$. But a collineation is completely determined by the

* This derivation of (2) and (3) applies to any desarguesian geometry, whereas the proof of (3) given by LEVI assumes the commutativity of multiplication. This is because LEVI defines an antiprojectivity of a line as a transformation which changes harmonic throws into harmonic throws. The two definitions have not been proved identical in cases where multiplication is non-commutative. DARBOUX's proof of the fundamental theorem of projective geometry amounts to showing that conditions (2), (3), (4) are satisfied in the case of ordinary number only by the identical transformation.

† Similarly if the marks of any $GF[p^\nu]$ included in the $GF[p^n]$ are required to be invariant, $\phi(x)=x^{(p^\nu)^l}$. From this consideration are derived the results of the last section of Dr. LEVI's paper.

permutations of the points of two fixed lines. There are thus $n-1$ antiprojectivities of the plane. Since the following are evidently $n-1$ distinct antiprojectivities they must therefore be all the antiprojectivities of the plane:

$$\frac{x_1'}{x_3'} = \frac{x_1^{p^l}}{x_3^{p^l}}, \qquad \frac{x_2'}{x_3'} = \frac{x_2^{p^l}}{x_3^{p^l}}. \tag{5}$$

The most general collineation is now evidently a product of (5) by the most general linear fractional transformation and that product is of type (1). It follows as a corollary that the order of the collineation group of $PG(k, p^n)$ is the order of the linear fractional group $LF(k+1, p^n)$ multiplied by n.

PRINCETON, N. J.,
November 13, 1906.

GEOMETRY IN WHICH THE SUM OF THE ANGLES OF EVERY TRIANGLE IS TWO RIGHT ANGLES*

BY

R. L. MOORE

In his article *Die Legendre'schen Sätze über die Winkelsumme im Dreieck*,† DEHN has shown what may be stated precisely by the aid of the following abbreviations:

H_{-a} denotes all those theorems of geometry which are logical consequences of HILBERT's axiom-groups I, II, III, IV, viz., his axioms of association, of order, of parallels, and of congruence respectively.

H_{-a-p} denotes all those theorems of geometry which are logical consequences of HILBERT's axiom groups, I, II, IV.

S denotes the assumption that the sum of the angles of every triangle is two right angles.‡

What DEHN has shown is that H_{-a} does not follow from H_{-a-p} and S. I wish to show that, nevertheless, if H_{-a-p} and S are true of a space, then either this space is a H_{-a} space or it is possible so to introduce ideal points that the space thus extended will be a H_{-a} space, these ideal points, moreover, being such that no one of them is between two points of the original space.

One might state this a little more suggestively, if less accurately, as follows: "While the parallel postulate, III, and thus all of that part of Hilbertian Geometry which follows without use of his 'Axiom of ARCHIMEDES' and 'Vollständigkeit Axiom,' can not indeed be proved from his other postulates I, II, IV with III replaced by S, nevertheless it can be shown that a space concerning which these postulates (I, II, IV, S) are valid must be, if not the whole, then at least a part, of a space in which III also is true."§

* Presented to the Society (Chicago) April 22, 1905, as part of a paper entitled *Sets of metrical hypotheses for geometry*. Received in revised form for publication December 26, 1906.

† Mathematische Annalen, vol. 53 (1900), pp. 404–439.

‡ This may be stated precisely as follows: If ABC is a triangle and C is between B and E then, in the angle ACE, there is a point D such that $\measuredangle ACD \equiv \measuredangle BAC$ and $\measuredangle DCE \equiv \measuredangle ABC$.

§ This result has an interesting connection with our spatial experience. Statements have been made to the effect that, since no human instruments, however delicate, can measure exactly enough to decide in every conceivable case whether the sum of the angles of a triangle is equal to two right angles (unless the difference between this sum and two right angles should exceed

Let us first establish a number of preliminary theorems. The theorems of H_{-a-p} are assumed and I shall take for granted, sometimes without giving exact reference, certain of those definitions and theorems of HILBERT's *Festschrift* which are founded on his axiom groups I, II, IV. Reference will also be made to HALSTED's *Rational Geometry* (R. G.) in case of theorems for which demonstrations without use of HILBERT's III are therein indicated.

For simplicity let us confine ourselves to one plane.

THEOREM I. *If two straight lines, a and b, are perpendicular to each other at a point O, then every straight line, c, in the plane ab, has a point in common with either a or b.*

*Dem.** This conclusion of course holds if c coincides with a or b. If it does not coincide with a or b, then it has (cf. HILBERT's Theorem 5 and definition following IV, 3 and Theorem 9) a point C within one of the four right angles into which a and b divide the plane ab. There exists a point A on that ray of this angle which belongs to a and a point B on that ray thereof which belongs to b. There is (cf. R. G., art. 47, theorem) a point D on c such that OD is perpendicular to c. On c there are two points, E and E', such that $DE' \equiv DE \equiv OD$ (cf. HILBERT's IV, 1 and IV, 2). Then (as may be seen with help of S), the angle DOE is one half a right angle and the angle $E'OD$ is one half a right angle. Accordingly the angle $E'OE$ is a right angle. Hence it is easily seen (cf. R. G., arts. 51 and 143) that E and E' can not both be within the right angle AOB. Therefore c, which contains both E and E', has a point within, and a point without, the angle AOB and thus must cut either OA or OB.

THEOREM 2. *If O lies between A and C and also between B and D, and if $\sphericalangle AOB$ is a right angle and $\sphericalangle ABD \equiv \sphericalangle ACD$, then there is a circle passing through A, B, C, and D.*†

Dem. There exist (cf. R. G., 47, 51 and HILBERT's IV, 4) perpendiculars

---

a certain minimum amount), it is therefore impossible to settle the question whether our space is Bolyai-Lobachewskian or Euclidean even though it be granted that it is one or the other.

Now the geometry which I wish to show to be a consequence of H_{-a-p} and S is one concerning whose resemblance to Euclidean space one may say even more. One may say (possibly a little roughly): "While, as DEHN has shown, it is indeed true that a space of which H_{-a-p} and S were true would not necessarily be strictly Hilbertian (in the sense of H_{-a}), neverthless no human being confined therein could ever distinguish it from a Hilbertian space even though he were supplied with instruments which could decide for him whether any two segments were exactly equal."

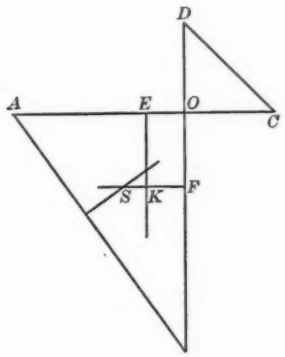
* "*Dem.*" means an *indication* of a demonstration. Likewise "by theorem" does not necessarily mean that this is the only theorem used in the demonstration. Similar statements may be made about the use, in this paper, of the words "therefore," "hence," etc.

Capital letters of the English alphabet designate points.

The sign "$=$" means "is, or are, identical with."

† This theorem gives a sufficient condition that three given points should lie on a circle, also a sufficient condition that 4 given points should lie on a circle.

LE and MF to AC and BD at their middle* points E and F respectively. As LE cannot meet BD (cf. R. G., art. 47), according to Theorem 1 it must meet FM at a point K. Now, by S and Theorem 1, the perpendicular to AB at its middle point must meet either KE or KF. Suppose it meets KF at a



point S. Then $SD \equiv SB \equiv SA$ (by definitions of a right angle and of a perpendicular, HILBERT's Theorem 10, and statement preceding Theorem 13). So there is a circle with S as center passing through A, B, and D. By R. G., art. 57, $\measuredangle SAD \equiv \measuredangle SDA$ and $\measuredangle SAB \equiv \measuredangle SBA$. It may then be seen, even without further use of S, that SAD and SAB are acute and hence SA is not perpendicular to AC and therefore, by R. G., art 138, that the circle which passes through A, B, D must have another point P in common with AC also that $\measuredangle DPA \equiv \measuredangle DBA$ (cf. S and R. G., art. 133); but this, in view of the hypothesis that $\measuredangle DBA \equiv \measuredangle DCA$, is impossible (cf. R. G., 66) unless P is C. The circle therefore through A, B, and D passes also through C. Similarly if the perpendicular to AB at its middle point met EK it would follow that there would be a circle through A, B, and C, and this circle would necessarily pass through D.

Convention. A point O, two straight lines perpendicular to each other at this point, a point I on one of these lines and points Q and S on the other, such that O is between Q and S, are selected once for all and considered as fixed throughout this discussion.

As an aid to the exhibition of an analytic geometry I wish to develop a calculus of those rays (half lines) which start from I towards the O-side of IK where IK is perpendicular to OI. These rays are called "leftward rays." If a leftward ray cuts QS it is called a "cutting ray." Small letters of the English alphabet are used to denote cutting rays.

*See R. G., articles 82, 64 and 66.

If M is a point of QS, $|IM|$ means IT where T is a point of the ray OQ such that $OM \equiv OT$.

The ray IO is designated by the symbol 0 (it is a "zero ray"). If I' is a point on the ray OQ such that $OI' \equiv OI$, then the ray II' is designated by the symbol 1 (it is a "unit ray").

Definition 1 (sum of two cutting rays). If L and M are two points (or the same point) on QS, ray IL + ray IM means ray IN where N is a point of QS such that $ON = OL + OM$ in the vector sense.

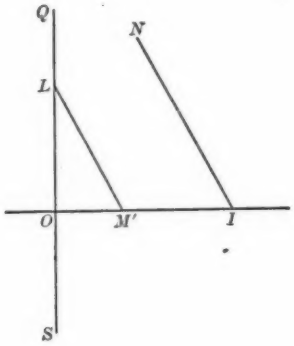
Definition 2. If IM and IN are two leftward rays, then "$IM > IN$" means that IN precedes IM in the sense $IS - IO - IQ$; and "$IM < IN$" means $IN > IM$.

THEOREM 3. *If α and β are two leftward rays, then of the three statements $\alpha > \beta$, $\alpha = \beta$, or $\alpha < \beta$ one and only one is true. If $\alpha < \beta$ and $\beta < \gamma$, then $\alpha < \gamma$.*

THEOREM 4. *The set of all cutting rays is a commutative group with respect to the operation of addition* $(+)$.

Dem. See Definition 1.

Definition 3 (quotient of two cutting rays). If L and M are points of the straight line QS and M is a point of ray OI such that $OM \equiv OM'$ and if IN is a ray such that $\measuredangle OIN \equiv \measuredangle OM'L$, if further IN lies on the Q- or the S-side of OI according as M and L are on the same side, or opposite sides, of O, then $IL/IM = IN$.



Definition 4 (product of two cutting rays), $ab = c$ if and only if $c/a = b$.

THEOREM 5. *If $a/b = c/b$, then $a = c$.*

THEOREM 6. *$b/1 = b$ and $b \times 0 = 0$.*

THEOREM 7. *If $a/b = c/d$, then $b/a = d/c$.*

THEOREM 8. *If A, B, C, D are points on QS and A and B are on the same side or opposite sides of O according as C and D are on the same side*

or on opposite sides of O, then the existence of four points A', B', C', D' on a circle and another point O' such that O' is between B' and C', and between A' and D', angle $A'O'B'$ is a right angle, $OA \equiv O'A'$, $OB \equiv O'B'$, $OC \equiv O'C'$, and $OD \equiv O'D'$, is a necessary and sufficient condition that $IA/IB = IC/ID$.

Dem. Use Definition 3 and Theorem 2.

THEOREM 9. *If $a/b = c/d$, then $a/c = b/d$.*

Dem. Use Theorem 8.

THEOREM 10. *If $a/b = c/d$ and $a/e = f/d$, then $b/e = f/c$.*

*Dem.** Consider three cases —

I. Suppose $|b| < |e|$. Evidently there exist an n such that $b/n = e/d$, and an x such that $b/x = e/c$ whence also $b/e = x/c$ (see Theorem 9). Now $a/f = e/d$ (see hypothesis and Theorem 9), so that $a/f = b/n$. Also $b/e = n/d$ (see Theorem 9), therefore $n/d = x/c$ and thus, by Theorem 9, $x/n = c/d$. But since, by hypothesis, $a/b = c/d$, we have $a/b = x/n$. But from $a/f = b/n$, by Theorem 9, $a/b = f/n$ and thus $x/n = f/n$; therefore, by Theorem 5, $x = f$. But as $b/e = x/c$, it follows finally that $b/e = f/c$.

II. Suppose $e < b$. Use similar reasoning to show that $e/b = c/f$ and thence conclude (by use of Theorem 7) that $b/e = f/c$.

III. Suppose $e = b$. In this case, by Theorem 5, $f = c$ and by Theorem 9 the theorem is true.

THEOREM 11. *If $ab = x$† and $a'b' = ab$, then $a/a' = b'/b$; and conversely, if $ab = x$ and $a/a' = b'/b$, then $a'b' = ab$.*

Dem. I. If $ab = x$ and $a'b' = ab$, then, according to Definition 4, $x/a = b$ and $x/a' = b'$. According to Theorems 6 and 7 $1/b' = a'/x$ and $1/b = a/x$. Hence, by Theorem 10, $b'/b = a/a'$.

II. If $ab = x$ and $a/a' = b'/b$, then, by Definition 4 and Theorem 2, $x/a = b/1$. Hence, by Theorem 7, $a/x = 1/b$ and therefore, by hypothesis and Theorems 10 and 6 and Definition 4, $x = a'b'$.

THEOREM 12. *If $ab = x$, then $ba = ab$.*

Dem. If $ab = x$, then, by Definition 4 and Theorem 6, $x/a = b/1$. Hence, by Theorems 9 and 6, $x/b = a$. Therefore, by Definition 4, $ba = x$. But since, by hypothesis, $ab = x$, $ba = ab$.

THEOREM 13. *If $ab = x$, $(ab)c = y$, and $bc = z$, then $a(bc) = (ab)c$.*

Dem. According to Theorem 11 this proposition will be established if it is proved that $ab/a = bc/c$. Now according to Definition 4 and Theorem 6 $ab/a = b/1$. Likewise $bc/c = b/1$. So $ab/a = bc/c$.

THEOREM 14. *If $ab = x$, $ac = y$, then $a(b + c) = ab + ac$.*

Dem. Use S in connection with 2nd figure on page 34 of HILBERT's *Grundlagen der Geometrie* (2nd edition).

* The suggestion of this demonstration I am unable at present to trace to its source.

† i. e., ab is a cutting ray.

THEOREM 15. *If $|a|<1$ then for every b, ab exists as a cutting ray. For every c there exists x such that $|cx|<1$.*

If $|a|<|b|$ and bx exists as a cutting ray, then $|ax|<|bx|$. Given d, e, k, $e\neq 0$, $k\neq 0$, then there exist d', e' such that $|d'|<|k|$, $|e'|<|k|$, and $d/e=d'/e'$.

THEOREM 16. *If $|a|$, $|b|$, $|c|$, $|d|$ are all less than 1, and if $a/b=e$ and $c/d=f$, then $a/b+c/d=(ad+bc)/bd$.*

Dem. According to hypothesis and Definition 4, $a=be$, $c=df$; and by hypothesis and Theorems 15, 12, 13, and 14, $ad+bc=(be)d+b(df)=d(be)+b(df)=(db)e+(bd)f=(bd)e+(bd)f=bd(e+f)$. This gives, by the use of Definition 4, $(ad+bc)/bd=e+f=a/b+c/d$.

THEOREM 17. *If $|a|$, $|b|$, $|c|$, $|d|$ are all less than 1, $a/b=e$, $c/d=f$ and $ef=y$, then $ac/bd=ef$.*

Dem. By hypothesis and Definition 4, $c=df$, $a=be$ and, by hypothesis and Theorem 15, $ac=(be)(df)$. Hence, by hypothesis, Theorems 15, 12, and 13, $ac=(bd)(ef)$, and from this, by Definition 4, $ac/bd=ef$.

THEOREM 18. *If $|a|$, $|b|$, $|c|$, $|d|$, $|a'|$, $|b'|$, $|c'|$, $|d'|$ are all less than 1, and if $a/b=a'/b'$, $c/d=c'/d'$, then $(ad+bc)/bd=(a'd'+b'c')/b'd'$.*

Dem. By hypothesis and Theorems 15 and 11, $ab'=ba'$ and $cd'=dc'$. Hence (see Theorem 15) $(ab')(dd')=(ba')(dd')$ and $(cd')(bb')=(dc')(bb')$. Hence, by Theorems 12, 13, and 14, $(ad)(b'd')+(bc)(b'd')=(bd)(a'd')+(bd)(b'c')$ and thus, by Theorems 12 and 14, $(ad+bc)b'd'=bd(a'd'+b'c')$; therefore by Theorem 11, $(ad+bc)/bd=(a'd'+b'c')/b'd'$.

THEOREM 19. *If $|a|$, $|b|$, $|c|$, $|d|$, $|a'|$, $|b'|$, $|c'|$, $|d'|$ are all less than 1 and $a/b=a'/b'$ and $c/d=c'/d'$, then $ac/bd=a'c'/b'd'$.*

Dem. According to Theorems 15 and 11, $ab'=ba'$, $cd'=dc'$, whence follows (see Theorem 15), $(ab')(cd')=(ba')(dc')$. Hence, by Theorems 15, 12, and 13, $(ac)(b'd')=(bd)(a'c')$ and thus, by Theorem 11, $ac/bd=a'c'/b'd'$.

Definition 5. If either there is no e such that $a/b=e$ or there is no f such that $c/d=f$, then $a/b\times c/d$ means $a'c'/b'd'$ and $a/b+c/d$ means $a'd'+b'c'/b'd'$ where a', b', c', d' are in absolute value less than 1, and $a/b=a'/b'$ and $c/d=c'/d'$.

THEOREM 20. *$a/b\times c/d=a'c'/b'd'$ and $a/b+c/d=a'd'+b'c'/b'd'$ where a', b', c', d' are all less than 1 in absolute value and $a'/b'=a/b$ and $c'/d'=c/d$.*

Dem. See Theorems 16 and 17 and Definition 5.

Lemma. If $a'b$, $a'c$ and $(a'b)c$ are cutting rays, then $a'b/a'c=b/c$.

Dem.
$$(a'b)c=(ba')c \quad \text{by hypothesis and Theorem 12,}$$
$$=b(a'c) \quad \text{by Theorem 13,}$$
$$=(a'c)b \quad \text{by Theorem 12.}$$

Hence, by Theorem 11, $a'b/a'c=b/c$.

THEOREM 21. *If b and d are different from 0 and a and c are any cutting rays, then there exist x and y $(y \neq 0)$ such that $a/b + x/y = c/d$. If $a/b + x/y = c/d$ and $a/b + w/z = c/d$, then $x/y = w/z$.*

Dem. Use Lemma and Theorems 15, 4, 20, 11.

THEOREM 22. *If a, b, and d are different from 0 and c is any cutting ray, then there exist x and y such that $a/b \times x/y = c/d$. If $a/b \times x/y = c/d$ and $a/b \times w/z = c/d$, then $x/y = w/z$.*

Dem. I. By Theorem 15 there exist a', b', c', d' all less than 1 in absolute value, such that $a/b = a'/b'$, $c/d = c'/d'$. By Theorems 15, 20, 13, 11, and Lemma, $a/b \times b'c'/a'd' = a(b'c')/b(a'd') = (ab')c'/(ba')d' = c'/d'$.

II. Suppose $a/b \times x/y = c/d$ and $a/b \times w/z = c/d$. By Theorem 15 there exist a', b', c', d', x', y', w', z', all less than 1 in absolute value, such that $a/b = a'/b'$, $c/d = c'/d'$, $x/y = x'/y'$, $w/z = w'/z'$. By hypothesis and Theorem 20, $a'x'/b'y' = c'/d'$ and $a'w'/b'z' = c'/d'$. Hence, by Theorems 15, 11, 12, 13, $(a'd')x' = (b'c')y'$ and $(a'd')w' = (b'c')z'$. Hence, by Theorems 12 and 11, $x'/y' = b'c'/a'd' = w'/z'$. Hence $x/y = w/z$.

Convention. Any Greek letter denotes a/b where a and b are cutting rays.

THEOREM 23. *The set of all α's is a number system for which Hilbert's Theorems* 1–16 *of* § 13 *of his Grundlagen der Geometrie hold true with respect to $+$, $\times$, and $>$ as defined in this paper.**

Dem. HILBERT'S Theorems 1–12 may be proved by use of my Theorems 20, 6, 4, 21, 22, 12, 15, 14, 13, and Lemma to Theorem 22. His Theorems 13–16 may be proved with the use of my Theorems 1–12 in connection with Definitions 1–5.

Definition 6. The length of a segment AB means the ray IT where T is a point on the ray OQ such that $AB \equiv OT$.

The length of AB is denoted by "$e(AB)$."

Lemma I. If in the triangles ABC, $A'B'C'$, $\sphericalangle A \equiv \sphericalangle A'$, $\sphericalangle B \equiv \sphericalangle B'$, $\sphericalangle C \equiv \sphericalangle C'$, then

$$\frac{e(AB)}{e(A'B')} = \frac{e(BC)}{e(B'C')} = \frac{e(CA)}{e(C'A')}.$$

Dem. Use Definitions 6 and 3, Theorem 23, etc., in connection with Theorem 22 of HILBERT'S *Grundlagen der Geometrie*.

Lemma II. If BAC is a right angle, then $e^2(AB) + e^2(AC) = e^2(BC)$.

Dem. There is (cf. R. G., Art. 47) a point D on BC such that AD is perpendicular to BC. D is between B and C. Otherwise (as may easily be seen, cf. R. G., Art. 66, etc.), either AD and AC or AD and AB could not meet on the A side of BC. According to S, $\sphericalangle ACB \equiv \sphericalangle BAD$ and

*See Definitions 1–5.

$\measuredangle CAD \equiv \measuredangle ABC$. It follows by Lemma I, Theorem 23 and Definitions 6 and 1, that

$$e^2(AB) + e^2(AC) = e(BD)e(BC) + e(CD)e(BC)$$
$$= \{e(BD) + e(CD)\}e(BC) = e^2(BC).$$

THEOREM 24. *Given α, β there exists γ such that $\gamma^2 = \alpha^2 + \beta^2$.*

Dem. This theorem is evidently true for the case in which α or $\beta = 0$. If α and β are different from 0, then by Theorems 6 and 15 there exist a, b, c, d, all less than 1 in absolute value, such that $\alpha = a/b$, $\beta = c/d$. By Theorem 23 $\alpha^2 + \beta^2 = [(ad)^2 + (bc)^2]/(bd)^2$. But, by Theorem 15 and Definition 6, ad and bc are lengths of segments. Hence there exists a right-angled triangle BAC such that the length of AB is ad and the length of AC is bc. According to Lemma 2, $e^2(BC) = (ad)^2 + (bc)^2$. Therefore, by Theorem 23, $\alpha^2 + \beta^2 = e^2(BC)/(bd)^2 = [e(BC)/bd]^2$.

Convention. Two straight lines OX and OY perpendicular to each other are selected as axes of coördinates. If P is any point and D and E are the feet of the perpendiculars from P to OX and OY respectively, then x_P means 0, $e(PE)$, or $-e(PE)$ according as P is on OY, on the X-side of OY, or on the non-X-side of OY, and y_P means 0, $e(PD)$, or $-e(PD)$, according as P is on OX, on the Y-side of OX, or on the non-Y-side of OX.

As a result of this convention we have the theorem:

THEOREM 25. *Every point P is represented by one and only one sensed pair of coördinates (x_P, y_P) and every sensed pair of cutting rays represents one and only one point.*

THEOREM 26. *Given a straight line AB, there exist α, β, γ (α, β not both 0) such that the x and y of every point on AB satisfy the equation $\alpha x + \beta y + \gamma = 0$.*

Dem. Use S, Theorems 1 and 23, lemma I, Definitions 6 and 1, also R. G., articles 47 and 49, and HILBERT's axiom IV, 4, Axiom IV, 1, Theorem 11, etc.

THEOREM 27. *If B is between A and C and $x_A \neq x_B$, then $x_A < x_B < x_C$ or $x_A > x_B > x_C$. If B is between A and C and $x_A = x_B$, then $y_A > y_B > y_C$ or $y_A < y_B < y_C$.*

Dem. Use Theorems 1 and 23, R. G., 47, and HILBERT's Theorem 5 et seq.

THEOREM 28. *The length $e(AB)$ equals $\sqrt{(x_A - x_B)^2 + (y_A - y_B)^2}$ and thus the relation $e(AB) = e(A'B')$ is a necessary and sufficient condition for the equality*

$$(x_A - x_B)^2 + (y_A - y_B)^2 = (x_{A'} - x_{B'})^2 + (y_{A'} - y_{B'})^2.$$

Dem. Use lemma 2 in connection with convention preceding Theorem 25.

THEOREM 29. *The existence of* $\alpha, \beta, \gamma, \delta$ *satisfying the relations*

$$x_{A'} = \frac{\alpha}{\sqrt{\alpha^2+\beta^2}} x_A - \frac{\beta}{\sqrt{\alpha^2+\beta^2}} y_A + \gamma,$$

$$x_{B'} = \frac{\alpha}{\sqrt{\alpha^2+\beta^2}} x_B - \frac{\beta}{\sqrt{\alpha^2+\beta^2}} y_B + \gamma,$$

$$y_{A'} = \frac{\beta}{\sqrt{\alpha^2+\beta^2}} x_A + \frac{\alpha}{\sqrt{\alpha^2+\beta^2}} y_A + \delta,$$

$$y_{B'} = \frac{\beta}{\sqrt{\alpha^2+\beta^2}} x_B + \frac{\alpha}{\sqrt{\alpha^2+\beta^2}} y_B + \delta,$$

is a necessary and sufficient condition for the congruence $AB \equiv A'B'$.

In view of Theorems 23–27 and Theorem 29 it may be easily seen that one could proceed in a manner similar to that indicated by HILBERT in § 9 of his *Grundlagen der Geometrie* and show that all of his axioms hold true of our geometry, were it not for one obstacle, namely that there is not necessarily a perfect one-to-one correspondence between the set of all points and the set of all sensed pairs of α's. If α is not a cutting ray then (α, β) does not correspond to a point in manner indicated in Theorem 25.

It is natural then to fill up this gap by means of the following definitions:

Definition 7. Every sensed pair (α, β) is called an ideal point and "ideal point" means such a sensed pair; α and β are called its coördinates.

Definition 8.* An ideal straight line means the set of all existent points, real and ideal, whose coördinates satisfy an equation of the form $\alpha x+\beta y+\gamma=0$.

Definition 9. If of the points A, B, C at least one is ideal then ABC means $x_A > x_B > x_C$ or $x_A < x_B < x_C$ unless $x_A = x_B$, in which case ABC means $y_A > y_B > y_C$ or $y_A < y_B < y_C$.

Definition 10. If one of the points A, A', B, B' is ideal then "$AB \equiv A'B'$" means there exist $\alpha, \beta, \gamma, \delta$ satisfying the relations

$$x_{A'} = \frac{\alpha}{\sqrt{\alpha^2+\beta^2}} x_A - \frac{\beta}{\sqrt{\alpha^2+\beta^2}} y_A + \gamma,$$

$$x_{B'} = \frac{\alpha}{\sqrt{\alpha^2+\beta^2}} x_B - \frac{\beta}{\sqrt{\alpha^2+\beta^2}} y_B + \gamma,$$

$$y_{A'} = \frac{\beta}{\sqrt{\alpha^2+\beta^2}} x_A + \frac{\alpha}{\sqrt{\alpha^2+\beta^2}} y_A + \delta,$$

$$y_{B'} = \frac{\beta}{\sqrt{\alpha^2+\beta^2}} x_B + \frac{\alpha}{\sqrt{\alpha^2+\beta^2}} y_B + \delta.$$

* In this definition and in what follows, small letters of the English alphabet are not always used, according to earlier convention, to designate cutting rays exclusively.

In view of Theorems 25, 26, 27, 29 and the corresponding Definitions 6, 7, 8, 9 it is evident that our real and ideal points and straight lines form a system which, with respect to order, congruence and association, is related to our set of α's as HILBERT's geometry of § 9 of his *Festschrift* is related to his number system of that paragraph, and moreover in view of Theorems 23 and 24 it is evident that the set of α's satisfy with respect to the operations of addition $(+)$, multiplication $(\times)$, and the relation $>$, sufficient conditions to enable us to proceed according to the method indicated by HILBERT to prove that our geometry satisfies his axioms of groups I–IV. It remains to be shown that no ideal point is between two real points. This may be proved by use of Definitions 7 and 1 and Theorem 23.

For helpful suggestions and criticisms the author desires to tender to Professor VEBLEN his grateful acknowledgment.

PRINCETON UNIVERSITY.

NON-DESARGUESIAN AND NON-PASCALIAN GEOMETRIES*†

BY

O. VEBLEN AND J. H. MACLAGAN-WEDDERBURN

§ 1. The object of this paper is to discuss certain non-pascalian and non-desarguesian geometries. Beside their intrinsic interest, these geometries are useful in forming independence proofs for systems of axioms for geometry.

In a previous paper‡ it has been shown how to construct all finite projective geometries of three or more dimensions and all two dimensional projective geometries in which the Desargues theorem (cf. third footnote) is valid. The present paper supplements the other by showing the existence of non-desarguesian finite geometries. It is to be noted that, since every finite linear associative algebra in which every number possesses an inverse is commutative, § therefore every non-pascalian finite geometry is necessarily non-desarguesian. Moreover, as has been shown by HESSENBERG,‖ every non-desarguesian geometry is also non-pascalian, whether finite or infinite.

A *projective plane geometry* is a set of elements called points, finite or infinite in number, subject to the following conditions:

1. If A and B are any two points, there is (α) one and (β) only one set of points called a line and containing both A and B.

2. If a and b are any two lines, there is (α) one and (β) only one¶ point contained in both a and b.

3. Each line contains at least three points.

* Presented to the Society (Chicago), April 22, 1905. Received for publication, December 6, 1906.

† As originally read, the paper called attention to the fact that HILBERT's example of a non-pascalian geometry in his *Festschrift* did not satisfy his order axioms. The error however is not repeated in later editions of his book. This paper also contained a discussion of the non-pascalian geometry obtained by letting the coördinates of a point be quaternions. As such geometries are discussed in the *Abstrakte Geometrie* of K. T. VAHLEN, Leipzig, 1905, this part of the paper is omitted.

‡ VEBLEN and BUSSEY, *Finite projective geometries*, Transactions, vol. 7 (1906), p. 271. The reader is referred to this paper for references to the literature and discussion of the general problem of finite geometries. Condition VII of that paper is the "Pascal Theorem" and the paragraph following VII contains the "Desargues Theorem."

§ J. H. MACLAGAN-WEDDERBURN, *A theorem on finite algebras*, Transactions, vol. 6 (1905), p. 349.

‖ G. HESSENBERG, *Beweis des Desarguesschen Satzes aus dem Pascalschen*, Mathematische Annalen, vol. 61 (1905), p. 161.

¶ The condition $2(\beta)$ is of course logically a consequence of $2(\alpha)$ and $1(\beta)$.

The geometry is non-desarguesian or non-pascalian if the theorems of DESARGUES or of PASCAL, respectively, fail to be valid.

If the number of points is finite it is at once evident that every line contains the same number of points and that if $n+1$ is the number of points on a line, the total number of points is n^2+n+1.

Conversely, if the total number of points is n^2+n+1 and the number of points on each line is $n+1$ and every two points lie on one and only one line, it follows that every two lines have one and only one point in common. For, suppose the two lines a, b had no point in common. A point of a would determine with b $n+1$ lines giving thus $n(n+1)+1$ points and, counting in the points of a, we should have at least n^2+2n points instead of n^2+n+1.

By a dual argument, if the number of lines is n^2+n+1 and the number of lines through each point is $n+1$ and every two lines have one and only one point in common, it follows that every two points determine one and only one line.

§ 2. Analytic geometries may be constructed so as to satisfy the given conditions if the coördinates are numbers of a number-system subject only to the following conditions:

(1) The numbers form a commutative group under the operation of addition: that is, for any two numbers a and b there is a unique number c such that $a+b=c$ and a unique number d such that $a+d=b=d+a$, and there is a number 0 such that, for every a, $a+0=a=0+a$; further, addition is associative and commutative,* i. e. $(a+b)+c=a+(b+c)$ and $b+a=a+b$.

(2) For any two numbers a and b there is a unique number c such that $ab=c$ and if $a\neq 0$ unique numbers d, d' such that $da=b$ and $ad'=b$; also $0a=0=a0$ for every a.

(3) Finally there must be some sort of a distributive law. We shall use the following: $c(a+b)=ca+cb$. We shall not use for the present† the other distributive law $(a+b)c=ac+bc$, nor the associative and commutative laws of multiplication, nor shall we assume that there exists an identity element with respect to multiplication.

A *point* is now defined as one of the systems of three coördinates

$$(\alpha) \qquad x,\ y,\ \phi,$$

$$(\beta) \qquad x,\ \phi,\ 0,$$

$$(\gamma) \qquad \phi,\ 0,\ 0,$$

* The commutativity of addition is not an independent postulate. Cf. H. HANKEL, *Theorie der complexen Zahlensysteme* (Leipzig, 1867), page 32; E. V. HUNTINGTON, Annals of Mathematics, vol. 8 (1906), page 25.

† Cf. however the last sentence of this §.

where ϕ is not equal to 0 and is the same for all points. There are $n^2 + n + 1$ points in all if there is a finite number, n, of elements in the number-system.

A *line* is defined as all points which satisfy an equation of one of the forms

$$x\psi + yb + zc = 0, \tag{1}$$

$$y\psi + zc = 0, \tag{2}$$

$$z\psi = 0, \tag{3}$$

where ψ is not equal to 0 and is the same for all lines. There are $n^2 + n + 1$ lines altogether in the finite case. Further there are, in that case, $n + 1$ points on a line. For (3) is satisfied by points of types (β) and (γ): in case $c = 0$ (2) is satisfied by the n points of type (α) in which $y = 0$ and by (γ): in case $c \neq 0$ in (2) if $z = 0$, then $y = 0$ and hence (γ) is one of the points, while if $z = \phi$, y must be equal to the solution of $y\psi = -\phi c$ and x may have any of n values. The argument is similar for an equation of type (1). A similar argument will show that in the finite case the number of lines through a point is $n + 1$.

To show that our points and lines form a projective plane geometry it is necessary to prove that any two lines have in common one and only one point. This requires the consideration of several cases.

(a) Two lines of type (1):

$$x\psi + yb + zc = 0,$$

$$x\psi + yb' + zc' = 0.$$

The coördinates of a point of intersection must satisfy the condition

$$y(b - b') + z(c - c') = 0.$$

In case $c = c'$, b cannot be the same as b' (else the lines would be identical) and hence $y = 0$; the lines then have in common only the point $(\alpha, 0, \phi)$ where α is the unique solution of $x\psi = -\phi c$. In case $c \neq c'$ and $b = b'$, we have $z = 0$, $y = \phi$ and x is the solution of $x\psi = -\phi b$. In case $c \neq c'$, $b \neq b'$ we have $z = \phi$ and hence y is the solution, say α, of $y(b - b') = -\phi(c - c')$ and x is the solution of $x\psi = -(\alpha b + \phi c)$. On account of the commutative law of addition the values of x, y, and z thus found must satisfy the given equations.

(b) Lines of types (1) and (2):

$$x\psi + yb + zc' = 0,$$

$$y\psi + zc' = 0.$$

The coördinate z cannot be zero, else y and x would also be zero. Hence $z = \phi$, y is the solution, α, of $y\psi = -\phi c'$, and x the solution of $x\psi = -(\alpha b + \phi c')$.

(c) Lines of types (1) and (3):

$$x\psi + yb + zc = 0,$$

$$z\psi = 0.$$

The coördinate y cannot be zero since then x and z would both be zero. Hence $y = \phi$ and x is the solution of $x\psi = -\phi b$.

(d) Two lines of type (2):

$$y\psi + zc = 0,$$

$$y\psi + zc' = 0.$$

Any solution must satisfy the condition $z(c - c') = 0$ or $z = 0$. Hence $y = 0$ and $x = \phi$.

(e) Lines of types (2) and (3):

$$y\psi + zc = 0,$$

$$z\psi = 0.$$

The only solution is $(\phi, 0, 0)$.

This, according to the last sentence of § 1, completes the proof that all finite number systems of the type described above lead to finite plane geometries. If the number system is not finite it is necessary to prove algebraically that every two points determine a unique line. If the other distributive law holds also or if multiplication is commutative this can be done by the analysis used above.

§ 3. In order to construct a particular non-desarguesian geometry it is necessary only to adduce a particular number-system. The first class of systems which we shall use is due to L. E. DICKSON.* It is such that the number of elements in each system is finite and such that multiplication is associative but not commutative and the distributive law holds only in the form $x(y + z) = xy + xz$. Each system possesses an identity which in the analysis above may be taken as the common value of ϕ and ψ. These systems are discussed on pages 5 to 22 of DICKSON's paper. A characteristic system of this type is the following. There are p^{2m} elements, (a, b), where a and b are marks of a field $GF[p^m]$, p being an odd prime. These elements obey the following laws of combination:

$$(a_1, b_1) + (a_2, b_2) = (a_1 + a_2, b_1 + b_2)$$

$$(a_1, b_1) \times (a_2, b_2) = (a_1 a_2 + \epsilon\nu b_1 b_2, b_1 a_2 + \epsilon a_1 b_2)$$

where ν is a fixed not-square in the $GF[p^m]$, and $\epsilon = \pm 1$ according as $a_1^2 - \nu b_1^2$ is a square or not-square, i. e., $\epsilon = (a_1^2 - \nu b_1^2)^{(p^m-1)/2}$.

*L. E. DICKSON, *On finite algebras*, Göttinger Nachrichten, 1905, pp. 358-394.

The simplest special case is where $p^m = 3$. The algebra may then be thought of as having two units 1 and j and composed of all numbers of the form $x + yj$, x and y being integers modulo 3 and $j^2 = -1$. The system* is thus composed of nine numbers $0, 1, 2, j, 2j, 1+j, 1+2j, 2+j, 2+2j$.

It is not difficult (as the reader may verify analytically) to find cases which show that the geometry founded on this number-system is non-desarguesian and non-pascalian. Instead of doing this, however, we shall now show how to get another geometry based on the same number-system. The new geometry is so constructed as to be capable of being explicitly exhibited in a compact form.

Since the number-system is associative and distributive [in the sense, $a(b+c) = ab + ac$] the point (x, y, z) may also be represented by $(\kappa x, \kappa y, \kappa z)$, $\kappa \neq 0$. With this understanding a linear transformation changes points into points and, if their coefficients are scalar, two transformations may be combined in the usual way. They do not in general change lines into lines.†

Consider now the following transformation of period 13:

$$A: \qquad x = x' + z', \qquad y = 2x', \qquad z = 2y' + 2z'.$$

Let $A_0, B_0, \cdots, G_0$ denote the points $(2,0,1), (2+2j, j, 1), (2+j, 2j, 1), (1+2j, 1+j, 1), (j, 2+2j, 1), (1+j, 1+2j, 1), (2j, 2+j, 1)$ and let $A_k, B_k, \cdots, G_k$ $(k<13)$ denote the points derived from these by transforming them by the kth power of A. In this notation, seven of the lines of our original geometry are:

$$x + y + z = 0: \; A_0A_1A_3A_9B_0C_0D_0E_0F_0G_0,$$
$$x + yj + z = 0: \; A_0B_1B_8D_3D_{11}E_2E_5E_6G_7G_9,$$
$$x + y(2j) + z = 0: \; A_0C_1C_8E_7E_9F_3F_{11}G_2G_5G_6,$$

* In this special case the distributive law has a very simple form; namely

$$(a+bj)j = -aj - bj^2 = b - aj$$

where neither a nor b is zero. Though the commutative law is not valid it is always true either that $xy = yx$ or that $xy = -yx$.

For the convenience of the reader in verifying computations, we subjoin the multiplication table:

	2	j	$2j$	$1+j$	$1+2j$	$2+j$	$2+2j$
2	1	$2j$	j	$2+2j$	$2+j$	$1+2j$	$1+j$
j	$2j$	2	1	$2+j$	$1+j$	$2+2j$	$1+2j$
$2j$	j	1	2	$1+2j$	$2+2j$	$1+j$	$2+j$
$1+j$	$2+2j$	$1+2j$	$2+j$	2	$2j$	j	1
$1+2j$	$2+j$	$2+2j$	$1+j$	j	2	1	$2j$
$2+j$	$1+2j$	$1+j$	$2+2j$	$2j$	1	2	j
$2+2j$	$1+j$	$2+j$	$1+2j$	1	j	$2j$	2

† For example, the transformation A changes $(1,0,0), (0,j,1) (j,j,1)$ which lie on $yj+z=0$ into $(1,2,0) (1,0,2+2j), (1+j, 2j, 2+2j)$ of which the first two, but not the third, lie on $x(1+j)+y(1+j)+z=0$.

$$x+y(1+j)+z=0\colon\ A_0B_7B_9D_1D_8F_2F_5F_6G_3G_{11},$$
$$x+y(2+2j)+z=0\colon\ A_0B_2B_5B_6C_3C_{11}E_1E_8F_7F_9,$$
$$x+y(1+2j)+z=0\colon\ A_0C_7C_9D_2D_5D_6E_3E_{11}F_1F_8,$$
$$x+y(2+j)+z=0\colon\ A_0B_3B_{11}C_2C_5C_6D_7D_9G_1G_8.$$

A and its powers give 84 other heptads of points which may be obtained by adding successively the integers $1, 2, \cdots, 12$ to the subscripts of the points of each of the above lines, the resultant subscripts being, of course, reduced modulo 13. *The ninety-one heptads so obtained are the lines of a new non-desarguesian and non-pascalian geometry.*

It is not difficult to verify *a posteriori* that this scheme does define a geometry. By trial we see that if two lines intersect in more than one point, then neither of these points can be of the form A_k or B_k. Further, the scheme is invariant under the following transitive substitution group:

$$1,\ (BDG)(CEF),\ (BGD)(CFE),$$
$$(BC)(DF)(EG),\ (BE)(CD)(FG),\ (BF)(CG)(DE).$$

Hence the same is true of the remaining letters. The other axioms may be tested in the same way. That the geometry is non-desarguesian and non-pascalian is proved by examples which are sufficiently indicated in the diagrams below.

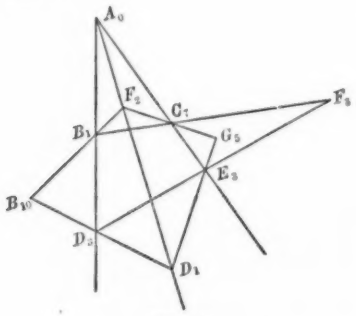


FIG. 1.

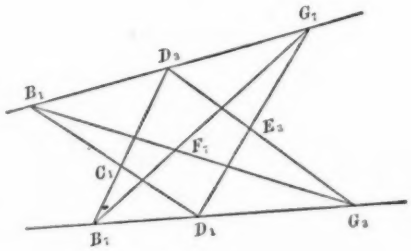


FIG. 2.

More interesting, however, than this tactical verification is the analytical proof that the 91 heptads derived by A and its powers from the lines B are the lines of a plane geometry. As in § 2, it is only necessary to show that any two heptads have one and only one point in common. It is a direct consequence of definition that a heptad is the set of points satisfying the equation,

$$x(a_{11}+a_{31})+y(a_{12}+a_{32})+z(a_{13}+a_{33})+(xa_{21}+ya_{22}+za_{23})\beta=0,$$

where β is any one of the seven numbers of the system and a_{rs} are the coefficients of any one of the thirteen powers of A. This equation may be abbreviated as follows:

$$xa + yb + zc + (xa' + yb' + zc')\beta = 0. \tag{1}$$

Consider the intersection of this with

$$x + y\alpha + z = 0. \tag{2}$$

In (1) and (2), α and β are the only coefficients which are not necessarily scalar. From (1) and (2) we deduce

$$y(2a\alpha + b) + z(2a + c) + \{y(2a'\alpha + b') + z(2a' + c')\}\beta = 0, \tag{3}$$

and shall now show that there is only one point whose coördinates satisfy (2) and (3) simultaneously. There are three cases.

(i) $2a' + c' = 0$. In this case (3) becomes:

$$y\{2a\alpha + b + (2a'\alpha + b')\beta\} + z(2a + c) = 0. \tag{4}$$

If the coefficients in (4) do not vanish identically, (2) and (4) have evidently one and only one point in common. On forming the powers* of the transformation A we find that $2a + c$ and $2a' + c'$ cannot vanish simultaneously except for the identical transformation, and in this case the vanishing of the coefficient of y leads to $\alpha = \beta$. Hence except in this trivial case (4) is not an identity.

(ii) $2a + c = 0$. In this case

$$y\{(2a\alpha + b)\beta^{-1} + 2a'\alpha + b'\} + z(2a' + c') = 0.$$

Hence as in (i) there is one and only one point common to (2) and (3).

(iii) $2a' + c' = (2a + c)d \neq 0$, where d is scalar. Then (3) may be written in the form

$$y\{2(a + 2a'd^{-1})\alpha + b + 2b'd^{-1}\} + \{y(2a'\alpha + b')d^{-1} + z(2a + c)\}(1 + d\beta) = 0,$$

or say

$$ye + (yf + zg)\gamma = 0.$$

This is equivalent to

$$y(e\gamma^{-1} + f) + zg = 0.$$

---

* The first twelve powers have the matrices

101	120	201	221	021	122	212	110	001	022	111	020
200	202	210	102	112	012	211	121	220	002	011	222
022	111	020	100	101	120	201	221	021	122	212	110

and the thirteenth power is the identical transformation.

Hence, since $g \neq 0$, (1) and (3) and therefore also (1) and (2) have one and only one point in common.

The point transformation A transforms lines of the new geometry into lines. Hence, if any two lines intersect in more than one point, the lines obtained by transforming these simultaneously by any power of A will also intersect in more than one point. But any pair of lines can, by transformation by a suitable power of A, be transformed into such a pair as has been considered above. Hence any two lines intersect in one and only one point.

§4. On pages 381–394, loc. cit.,* L. E. DICKSON has given another set of algebras which are available for geometrical purposes. The following multiplication table (cf. (68), loc. cit., p. 394) defines an algebra which satisfies the conditions of §2 together with the commutative law of multiplication and has an identity element. It therefore leads to a geometry.

	i	j
i	j	$b+\beta i$
j	$b+\beta i$	$-\beta^2-8bi-2\beta j$

Any number of the algebra is of the form $a+di+ej$, where a, d and e are any marks of any field F, not having the modulus 2, in which $x^2=b+\beta x$ is irreducible. In such an algebra an equation of the form $ax=b$ has a unique solution if $a \neq 0$.

If $b=2$, $\beta=0$ and F is the field of rational numbers this table becomes

	i	j
i	j	2
j	2	$-16i$.

Points and lines are defined as above, but as multiplication is commutative the equation of a line is more conveniently written in the form $ax+by+cz=0$.

That the geometry is non-desarguesian is shown by the following example. Consider the six points $A=(0,0,1)$, $A'=(0,i,1)$, $B=(i,i,1)$, $B'=(i,j,1)$, $C=(j,0,1)$, $C'=(j,i,1)$. These points lie in pairs on the lines $x=0$, $x-iz=0$, and $x-jz=0$, which meet in the point $(0,1,0)$. The equations of the sides of the triangles ABC and $A'B'C'$ are as follows:

$$AC\colon y=0, \qquad A'C'\colon y-iz=0,$$

$$AB\colon x-y=0, \qquad A'B'\colon x+(1+i+j)y/17-(2+i+j)z/17=0,$$

$$BC\colon x-(1-i)y-jz=0, \qquad B'C'\colon x+y-(i+j)z=0.$$

*See also L. E. DICKSON, *Linear algebras in which division is always uniquely possible*, these Transactions, vol. 7 (1906), p. 371.

If D, E and F are the intersections of AC with $A'C'$, AB with $A'B'$, and BC with $B'C'$, the coördinates of these points are: $D=(1,0,0)$, $E=(x=y=8(17/8+2i+j)/177, 1)$ and $F=[(-1+i+2j)/3, (1+2i+j)/3, 1]$. The equation of the line DE is $y-8(17/8+2i+j)z/177=0$ and evidently does not contain F. Hence D, E and F are not collinear and the geometry is non-desarguesian.

In the same way by taking the six points $A'=(1,0,1)$, $B'=(i,0,1)$, $C'=(i+j,0,1)$, $A=(1,1,0)$, $B=(i,1,0)$, $C=(j,1,0)$, it can be shown that the geometry is non-pascalian.

The *finite* number-system of the present type with the smallest number of elements is obtained by letting F be the $GF[3]$ and letting $b=1$ and $\beta=1$. This however leads to a geometry with 28 points on a line and $(27)^2+27+1=757$ points in all and has not been exhibited explicitly.

§ 5. The number systems discussed above all have identity elements. The presence of an identity is not necessary for the analysis of § 2, and examples of number systems that lead to geometries, though lacking the identity, are not hard to find. Two simple cases are constructed as follows:

Let e_1 and e_2 be two independent units. The systems under consideration consist of all numbers of the form ae_1+be_2 where a and b are integers reduced modulo 2. Addition is defined by the equation

$$(ae_1+be_2)+(ce_1+de_2)=(a+c)e_1+(b+d)e_2.$$

The two multiplication tables are

(i)

	e_1	e_2	e_3
e_1	e_3	e_2	e_1
e_2	e_2	e_1	e_3
e_3	e_1	e_3	e_2

(ii)

	e_1	e_2	e_3
e_1	e_1	e_2	e_3
e_2	e_3	e_1	e_2
e_3	e_2	e_3	e_1

where $e_3=e_1+e_2$. In (i) multiplication is commutative and in both systems the two-sided distributive law holds. Each system gives rise to a geometry with five points on a line. Both geometries however turn out to be identical with $PG(2,2^2)$ which is based on the Galois field, $GF(2^2)$. This suggests the following theorem: *There is only one projective plane geometry with five points on a line.*

To prove this consider the following Pappus configuration in which O, A, B, C, A_1, B_1, C_1 are distinct. Consider the lines DE, EF, FD. These lines obviously cannot intersect the line OA in any of the points A, B, C; hence as there are only two other points on OA, D, E, and F must be collinear. The geometry is therefore pascalian. DESARGUES's theorem may be proved in the

same way or it may be deduced from PASCAL's theorem. Now there is only one pascalian geometry for a given finite number of points on a line,* therefore there is only one geometry with five points on a line.

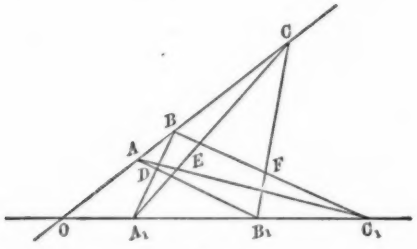


FIG. 3.

Thus a non-desarguesian geometry must have at least six points to the line. A number system which furnishes a geometry with this number of points is the following. Denote the elements by $0, 1, 2, 3, 4$, and let addition be identical with ordinary addition modulo 5. Let multiplication be defined by the table:

	1	2	3	4
1	1	2	3	4
2	4	3	2	1
3	3	1	4	2
4	2	4	1	3

The element, 1, is a left-hand identity. All the conditions of § 2 are satisfied and we may conveniently take $\phi = \psi = 1$. The resulting geometry however turns out to be $PG(2, 5)$. The problem of determining a non-desarguesian geometry of minimum number of points to the line remains unsolved.

* Cf. these transactions, vol. 7 (1906), p. 247 and § 1 above.

[Note added April 22, 1907. Dr. C. R. MACINNES has found by a tactical investigation that the only plane projective geometry of six points to a line is the $PG(2, 5)$ and that there is none at all with seven points to a line. The latter result has also been found by SAFFORD, American Mathematical Monthly, vol. 14 (1907), p. 84.]

MODULAR THEORY OF GROUP-MATRICES*

BY

LEONARD EUGENE DICKSON

1. The importance of the concept group-matrix in the theory of finite groups was recognized by DEDEKIND as early as 1880. The development of a general theory of group-matrices is due, however, to FROBENIUS, (Berliner Sitzungsberichte, from 1896 to the present). In particular, FROBENIUS has applied the theory to the representation of a finite group as a (non-modular) linear group. Since linear congruence groups are of importance in the theory of groups, particularly in questions of isomorphisms, and play a fundamental rôle in the applications of groups to the theory of functions and to geometry, the study of the representations of a finite group as a linear congruence group is of decided importance.

It is here proved that, if p^π is the highest power of the prime p dividing the order g of a group G, the group-matrix of G can be transformed, by a matrix whose elements are integers taken modulo p, into a compound matrix in which the submatrices to the right of the main diagonal have zero elements throughout, while the p^π submatrices in the diagonal are identical. Let D denote one of the diagonal submatrices, so that D is a square matrix of order g/p^π. Then the group-determinant Δ of G is congruent to $|D|^{p^\pi}$ modulo p. This result is in marked contrast to the non-modular theory, in which each algebraically irreducible factor of Δ enters to a power exactly equal to its degree.

It is shown in § 8 that the group-matrices of all groups of order p^π can be reduced to their canonical form modulo p by one and the same transformation.

An interesting theorem on group-characters is obtained in § 11.

Just as simplicity is attained in the algebraic theory only when certain algebraic irrationalities are introduced to permit of the complete factoring of the group-determinant into algebraically irreducible factors, so corresponding simplicity in the modular theory can be attained only by the use of Galois imaginaries (roots of irreducible congruences) in order to normalize completely the diagonal matrices D (§ 10). It is therefore proposed to take as the field of reference the field F_p defined as the aggregate of the Galois fields $GF[p^n]$,

* To be presented to the Society, September 5, 1907. Received for publication May 10, 1907.

$n = 1, 2, 3, \cdots$. Every equation with coefficients in F_p is completely solvable within F_p.

When the modulus p does not divide the order of G, the fundamental theorems of the algebraic theory are also true in the field F_p, as was first pointed out in the writer's paper in these Transactions, vol. 3 (1902), pp. 285–301. For the same case, the exposition of the algebraic theory by SCHUR (Berliner Sitzungsberichte, 1905, pp. 406–432) is valid in the field F_p.

When the modulus divides the order of G, the problem presents marked contrasts to the algebraic theory. It is to this outstanding case that the present paper is directed, as also the companion paper, "Modular theory of group-characters" (Bulletin of the American Mathematical Society, July, 1907).

2. Let G be a group of finite order g; let H be a subgroup,

$$H: \qquad s_1 = I, s_2, \cdots, s_h,$$

of order h and index $q = g/h$. Let $e_1 = I, e_2, \cdots, e_q$ be right-hand extenders of H to G, so that

$$G = H + He_2 + \cdots + He_q.$$

Form the left-hand multiplication-table of G with the operations

$$I, e_2, \cdots, e_q;\ s_2, s_2e_2, \cdots, s_2e_q;\ s_3, s_3e_2, \cdots, s_3e_q;\ \cdots$$

in the first row, and their inverses, in this order, in the first column. The body of the table is a compound matrix M with h^2 matrices $M_{i,j}$ as its elements. The matrix in the ith row of matrices and jth column of matrices is

$$M_{i,j} = M_k, \qquad \text{if} \qquad s_i^{-1}s_j = s_k. \tag{1}$$

We have

$$M_k = (e_r^{-1}s_ke_c) = \begin{pmatrix} s_k & s_ke_2 & \cdots & s_ke_q \\ e_2^{-1}s_k & e_2^{-1}s_ke_2 & \cdots & e_2^{-1}s_ke_q \\ \cdots & \cdots & \cdots & \cdots \\ e_q^{-1}s_k & e_q^{-1}s_ke_2 & \cdots & e_q^{-1}s_ke_q \end{pmatrix}. \tag{2}$$

If H is a subgroup of order h of a group G, the body of the multiplication-table of G may be exhibited as a compound matrix whose h^2 submatrices M_k have the same relative arrangement as the elements s_k in the multiplication-table $(s_i^{-1}s_j)$ of H.

To the h elements s_i of a group H we make correspond h independent variables x_{s_i}. Then the group-matrix of H is

$$(x_{s_i^{-1}s_j}) \qquad (i, j = 1, \cdots, h). \tag{3}$$

3. Let p^π be the highest power of p dividing the order $p^\pi q$ of a group G, and let H be a Sylow subgroup of order p^π.

It will be shown in § 5 that there exists a matrix (b_{ij}) which transforms the group-matrix (3) of H into a matrix whose diagonal elements are congruent, modulo p, to $\sum x_{s_i}$, and whose elements to the right of the diagonal are congruent to zero. Then, if I denotes the unit matrix of order q, the compound matrix $(b_{ij}I)$ transforms the group-matrix of G, arranged as in § 2, into a compound matrix, each of the p^π submatrices in the main diagonal being congruent to

$$\left(\sum_{k=1}^{p^\pi} x_{e_r^{-1}s_k e_c}\right) \qquad (r, c = 1, \cdots, q), \tag{4}$$

while each submatrix to the right of the main diagonal has all its elements congruent to zero modulo p. Hence we may state the following

THEOREM. *If p^π is the highest power of p dividing the order of a group G, the group-determinant of G is congruent modulo p to D^{p^π}, where D is the determinant of the matrix* (4).

Consider the linear transformation T on ξ_i $(i = 0, 1, \cdots, p^\pi - 1)$ whose matrix is the group matrix of H_{p^π} with a convenient order for the elements x_{s_i} in the first row (§ 5). When T is expressed in terms of the new variables

$$\eta_i = \sum_{j=i}^{p^\pi - 1} \binom{j}{i} \xi_j \qquad (i = 0, 1, \cdots, p^\pi - 1), \tag{5}$$

the coefficients being binomial coefficients, it takes the canonical form (modulo p) given above. The corresponding normalization of the transformation whose matrix is the group-matrix of G is accomplished by the introduction of the new variables

$$\eta_{iq+s} = \sum_{j=i}^{p^\pi - 1} \binom{j}{i} \xi_{jq+s} \quad (i = 0, 1, \cdots, p^\pi - 1;\ s = 1, \cdots, q). \tag{6}$$

4. Matrix (4) has the important property that, in the terminology of FROBENIUS, it is *a matrix belonging to the group G*. Removing the restriction that H is a Sylow subgroup, we obtain the matrix

$$X = \left(\sum_{k=1}^{h} x_{e_r^{-1}s_k e_c}\right) \qquad (r, c = 1, \cdots, q). \tag{7}$$

We prove that matrix X belongs to the group G, viz., that

$$XY = Z \quad \text{if} \quad \sum_R x_R y_{R^{-1}A} = z_A \qquad (R \text{ ranging over } G). \tag{8}$$

The element in the rth row and cth column of XY is

$$\sum_{j=1}^{q}\left(\sum_{m=1}^{h} x_{e_r^{-1}s_m e_j}\right)\left(\sum_{l=1}^{h} y_{e_j^{-1}s_l e_c}\right) = \sum_{k=1}^{h}\sum_R x_R y_{R^{-1}e_r^{-1}s_l e_c} = \sum_{k=1}^{h} z_{e_r^{-1}s_l e_c},$$

where $R = e_r^{-1}s_m e_j$, $s_k = s_m s_l$.

A second proof that matrix X belongs to the group G results incidentally from the following discussion which gives a (partial) normalization of the group-matrix of G by an algebraic transformation, in contrast to the (more complete) normalization in § 3 by a modular transformation. The transformation on $\xi_1, \cdots, \xi_{qh}$ whose matrix is the group-matrix of G(§ 2) is

$$\xi'_{(i-1)q+r} = \sum_{j,t} x_{e_r^{-1}s_i^{-1}s_j e_t} \xi_{(j-1)q+t} \qquad (i,j=1,\cdots,h;\ r,t=1,\cdots,q). \tag{9}$$

Consider the functions analogous to (6) with $i=0$:

$$\eta_s = \sum_{j=1}^{h} \xi_{(j-1)q+s} \qquad (s=1,\cdots,q).$$

Then

$$\eta'_r = \sum_{j,t}\left(\sum_{i=1}^{h} x_{e_r^{-1}s_i^{-1}s_j e_t}\right)\xi_{(j-1)q+t} = \sum_{t}\left(\sum_{k=1}^{h} x_{e_r^{-1}s_k e_t}\right)\xi_{(j-1)q+t} = \sum_{t=1}^{q}\left(\sum_{k=1}^{h} x_{e_r^{-1}s_k e_t}\right)\eta_t.$$

Hence the variables $\eta_1, \cdots, \eta_q$ undergo a transformation whose matrix is (7). In particular, $|X|$ is an algebraic factor of the group determinant of G.

5. THEOREM. *If G is a group of order p^π, p a prime, with the operations g_i, the group-matrix of G can be transformed* into one whose diagonal elements are congruent to $\sum x_{g_i}$ modulo p, while the elements to the right of the main diagonal are congruent to zero.*

The proof is by induction from $\pi-1$ to π. It proved in § 7 for $\pi=1$.

To apply § 2, let H be a subgroup of order $p^{\pi-1}$. Then $q=p$, and we may set $e_i = e^{i-1}$ $(i=2,\cdots,p)$. By the hypothesis for the induction, the group matrix (x_{s_k}) of H can be transformed into a matrix whose diagonal elements are congruent to $\sum x_{s_k}$ modulo p, while those to the right of the main diagonal are congruent to zero. Let the transforming matrix be (a_{ij}), $i,j=1,\cdots,p^{\pi-1}$. Then, if I is the unit matrix of order p, the compound matrix $(a_{ij}I)$ will transform the group-matrix (x_{g_i}) of G_{p^π} into a compound matrix, each of whose diagonal matrices D_i is (7) for $h=p^{\pi-1}$, while the elements of the matrices to the right of the diagonal are zero. Now $H_{p^{\pi-1}}$ is invariant in G_{p^π}. Hence $e^{-1}s_k e$ ranges over the operations of H where s_k does, so that

$$\sum_{k=1}^{p^{\pi-1}} x_{e_r^{-1}s_k e_c} = \sum x_{e^{-(r-1)}s_k e^{c-1}} = \sum x_{s_k e^{c-r}}.$$

We introduce the abbreviation

$$\sigma_t = \sum_{k=1}^{p^{\pi-1}} x_{s_k t}. \tag{10}$$

Since e^p belongs to H, we have

$$\sigma_{te^p} = \sigma_t. \tag{11}$$

* The transforming matrix is given in § 8.

Hence each diagonal matrix D_i is the cyclic matrix

$$(12) \qquad (\sigma_{c^{c-r}}) = \begin{pmatrix} \sigma_I & \sigma_c & \sigma_{c^2} & \cdots & \sigma_{c^{p-1}} \\ \sigma_{c^{p-1}} & \sigma_I & \sigma_c & \cdots & \sigma_{c^{p-2}} \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ \sigma_c & \sigma_{c^2} & \sigma_{c^3} & \cdots & \sigma_I \end{pmatrix}.$$

By § 7, this matrix of the cyclic group of order p can be transformed into a matrix whose diagonal elements are congruent to

$$\sum_{i=0}^{p-1} \sigma_{c^i} = \sum_{j=1}^{p^\pi} x_{\theta_j},$$

modulo p, while the elements to right of the diagonal are $\equiv 0$.

6. It remains to normalize the group-matrix of the cyclic group of order p by a transformation modulo p. Without increasing the difficulty of the problem, we treat the cyclic group of order p^π. We shall need two algebraic lemmas.

The transformation, with binomial coefficients,

$$(13) \qquad \eta_i = \sum_{j=i}^{g-1} \binom{j}{i} \xi_j \qquad (i = 0, 1, \cdots, g-1),$$

has, algebraically, the following inverse:

$$(14) \qquad \xi_i = \sum_{j=i}^{g-1} (-1)^{i+j} \binom{j}{i} \eta_j \qquad (i = 0, 1, \cdots, g-1).$$

This will be the case if, and only if,

$$(15) \qquad \sum_{j=i}^{l} (-1)^{l+j} \binom{l}{j} \binom{j}{i} = \delta_{il} \qquad (\delta_{il} = 1, \ \delta_{il} = 0 \text{ if } i \neq l).$$

This well known formula (cf. NETTO, *Combinatorik*, p. 255, formula 43) is a special case of the following one, which is needed later:

$$(16) \qquad E_{mrs} \equiv \sum_{i=0}^{r} (-1)^i \binom{r}{i} \binom{i+m}{s} = \begin{cases} 0 & (r > s) \\ (-1)^r \binom{m}{s-r} & (r \leqq s). \end{cases}$$

To evaluate E, we apply

$$\binom{r}{i} = \binom{r-1}{i} + \binom{r-1}{i-1}$$

and obtain two sums P and Q. Making a similar replacement for $\binom{i+m}{s}$ in Q, we get

$$Q = -\sum_{i-1=0}^{r-1} (-1)^{i-1} \binom{r-1}{i-1} \binom{i-1+m}{s} - \sum_{i-1=0}^{r-1} (-1)^{i-1} \binom{r-1}{i-1} \binom{i-1+m}{s-1}$$

$$= -P - E_{m,\, r-1,\, s-1}.$$

Hence we have the recursion formula

$$E_{mrs} = -E_{m,\,r-1,\,s-1}.$$

Thus for $r \leqq s$, $E_{mrs} = (-1)^r E_{m0s-r}$ and (16) follows. For $r - s = \rho > 0$,

$$E_{mrs} = (-1)^r E_{m\rho 0} = (-1)^r \sum_{i=0}^{\rho} (-1)^i \binom{\rho}{i} = (-1)^r (1-1)^\rho = 0.$$

For the case $m \leqq s$, formula (16) may be obtained from formula (41), (NETTO, loc. cit., p. 255), by replacing m, n, q, s by r, s, $s-m$, $i-s+m$, respectively.

7. THEOREM. *The group-matrix of the cyclic group of order p^π can be transformed into a matrix, given by* (22), *having congruent elements, modulo p, in every parallel to the main diagonal, the elements to the right of the latter being congruent to zero.*

The group-matrix of the cyclic group of order g is

$$A = \begin{pmatrix} \alpha_0 & \alpha_1 & \alpha_2 & \cdots & \alpha_{g-2} & \alpha_{g-1} \\ \alpha_{g-1} & \alpha_0 & \alpha_1 & \cdots & \alpha_{g-3} & \alpha_{g-2} \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \alpha_1 & \alpha_2 & \alpha_3 & \cdots & \alpha_{g-1} & \alpha_0 \end{pmatrix}. \tag{17}$$

Let $\alpha_{-k} = \alpha_{g-k}$. Then A is the matrix of the transformation

$$\xi_i' = \sum_{j=0}^{g-1} \alpha_{j-i}\, \xi_j \qquad (i = 0, 1, \cdots, g-1). \tag{18}$$

Introducing the variables η by means of (13) and (14), we get

$$\eta_s' = \sum_{r=0}^{g-1} \gamma_{sr} \eta_r, \qquad \gamma_{sr} = (-1)^r \sum_{i=0}^{r} \sum_{j=s}^{g-1} (-1)^i \binom{r}{i} \binom{j}{s} \alpha_{i-j}. \tag{19}$$

The value of γ_{sr} will be determined modulo p, for the case $g = p^\pi$.

For $0 \leqq m \leqq g-1$, the coefficient of α_{-m} in $(-1)^r \gamma_{sr}$ is $c + c'$, where c is the sum of the terms with $j = i + m$, c' the sum with $j = i + m - g$.

Let first $j = i + m$. The minimum value of i is the greater of 0 and $s - m$. But, for $m < s$ and $i < s - m$, $\binom{j}{s} = 0$. Hence in c we may allow i to begin with the value zero. The maximum value of i is the lesser of r and $g - 1 - m$. Hence if $m < g - r$, c is given by (16). Next, let $m \geqq g - r$, so that $i \leqq g - 1 - m$. Now c will be unaltered modulo p if we allow i to take the additional values $g - m + t$ $(t = 0, 1, \cdots, s-1)$, since

$$\binom{p^\pi + t}{s} \equiv 0 \pmod{p} \qquad (t < s < p^\pi).$$

Indeed, there is no term z^s in

$$(1 + z^{p^\pi})(1+z)^t \equiv (1+z)^{p^\pi + t} \pmod{p}. \tag{20}$$

Thus, if $g-r \leqq m < g-r+s$, then $g-m+s-1 \geqq r$ and c is congruent to (16). Finally, for $m \geqq g-r+s$, $c+c'$ is congruent to (16) in view of (21).

Let next $j = i+m-g$. Then i ranges from $g-m+s$ to r, so that terms c' occur only for $m \geqq g-r+s$. For $s < p^\pi = g$, $t < g$,

$$\binom{t}{s} \equiv \binom{g+t}{s}, \qquad \binom{i+m-g}{s} \equiv \binom{i+m}{s} \pmod{p},$$

by (20). Hence $c' = 0$ if $m < g-r+s$; while for $m \geqq g-r+s$,

$$c' \equiv \sum_{i=g-m+s}^{r} (-1)^i \binom{r}{i}\binom{i+m}{s} \pmod{p}. \tag{21}$$

Therefore in every case the coefficient of α_{-m} in $(-1)^r \gamma_{sr}$ is congruent modulo p to the sum (16). Thus

$$\gamma_{rs} \equiv 0 \text{ (if } r > s), \qquad \gamma_{sr} \equiv \sum_{m=0}^{g-1} \binom{m}{s-r} \alpha_{-m} \text{ (if } r \leqq s).$$

Hence $\gamma_{sr} \equiv \gamma_{s-r\,0}$ $(r \leqq s)$. Set $a_t = \gamma_{t0}$. Then (19) becomes

$$\eta_s' = \sum_{t=0}^{s} a_t \eta_{s-t}, \qquad a_t \equiv \sum_{m=0}^{g-1} \binom{m}{t} \alpha_{-m} \pmod{p}. \tag{22}$$

8. THEOREM. *The group-matrices of all groups of order p^π, with a suitable order for the elements of the first row, can be transformed simultaneously into their canonical forms modulo p by the same transforming matrix.*

Employing the notations of § 5, we prove that the transformation T on the variables ξ_i, whose matrix is the group-matrix (y) of a given group G of order p^π, can be reduced to its canonical form modulo p by the introduction of the new variables (5), viz., that transformation (5) transforms T into its canonical form. Then the matrix of (14), i. e., the inverse of the matrix of (5), will transform the group-matrix (y) into its canonical form modulo p. The proof is by induction from $\pi-1$ to π. We therefore assume that the transformation whose matrix is the group-matrix (x) of the subgroup H of order $p^{\pi-1}$ can be reduced to its canonical form by the introduction of the new variables

$$\eta_i = \sum_{j=i}^{a} \binom{j}{i} \xi_j \qquad (i = 0, 1, \cdots, a = p^{\pi-1}-1). \tag{23}$$

The matrix (a_{ij}) of § 5 is thus the inverse of the matrix of (23). Hence $(a_{ij}I)$ is the inverse of the matrix of

$$\eta_{ip+s} = \sum_{j=i}^{a} \binom{j}{i} \xi_{jp+s} \quad (i = 0, 1, \cdots, a;\ s = 0, 1, \cdots, p-1). \tag{24}$$

It remains to normalize the diagonal matrices D_k $(k = 0, 1, \cdots, a)$, each of

which is of the form (12). Hence by § 6 we introduce the new variables

$$\zeta_{kp+i} = \sum_{j=1}^{p-1} \binom{j}{i} \eta_{kp+j} \quad (k=0, 1, \cdots, a;\ i=0, 1, \cdots, p-1). \tag{25}$$

Eliminating the η's from (24) and (25), we get

$$\zeta_{lp+r} = \sum_{s=r}^{p-1} \sum_{m=l}^{a} \binom{s}{r} \binom{m}{l} \xi_{mp+s} \quad (l=0, 1, \cdots, a;\ r=0, 1, \cdots, p-1). \tag{26}$$

To prove that (26) is equivalent modulo p to transformation (5), set

$$i = lp + r, \qquad j = mp + s.$$

In view of the limits of the summation indices in (26), we have

$$0 \leqq r \leqq s \leqq p-1, \qquad 0 \leqq l \leqq m \leqq a \qquad (a = p^{\pi-1}-1).$$

It therefore follows from the writer's Dissertation (Annals of Mathematics, ser. 1, vol. 11 (1896–7), pp. 75, 76), that

$$\binom{j}{i} \equiv \binom{s}{r}\binom{m}{l} \pmod{p}. \tag{27}$$

The induction is therefore complete.

9. We have now established the lemmas employed in the proof of the general theorem of § 3. The problem of the ultimate canonical form modulo p of the group-matrix of a given group G of order $p^\pi q$ is therefore reduced to the problem of normalizing the diagonal matrices (4). One step of this normalization is readily effected in § 10, the resulting matrix (34) having the desirable property that the elements of the non-diagonal submatrices are all zero. In particular, this normalization is complete if a Sylow subgroup of order p^π is invariant in G.

10. Let H_{p^π} be invariant in $K_{p^\pi m}$, but in no larger subgroup of the given group $G_{p^\pi q}$. Set

$$K = H + He_2 + \cdots + He_m, \qquad G = K + Kf_2 + \cdots + Kf_n \quad (n = q/m). \tag{28}$$

Then (4) may be exhibited as a compound matrix (E_{ij}) whose n^2 elements E_{ij} are square matrices of order m:

$$E_{ij} = \left(\sum_{k=1}^{p^\pi} x_{f_i^{-1} e_r^{-1} s_k e_c f_j} \right) \qquad (r, c = 1, \cdots, m).$$

For r fixed, $e_r^{-1} s_k$ and $s_k e_r^{-1}$ range over the same set of elements. Set

$$\sigma_t^{ij} = \sum_{k=1}^{p^\pi} x_{f_i^{-1} s_k t f_j} \qquad (t = I, e_2, \cdots, e_m), \tag{29}$$

the m functions with i and j fixed being independent. Hence

$$E_{ij} = (\sigma^{ij}_{e_r^{-1}e_c}) \qquad (r, c = 1, \cdots, m). \tag{30}$$

Now $e_r^{-1}e_c$ may be given the form se_v with s in H. Then

$$\sigma^{ij}_{e_r^{-1}e_c} = \sigma^{ij}_{e_v}, \qquad (He_r)^{-1}(He_c) = He_v. \tag{31}$$

Hence, *for each i and j, E_{ij} is the group-matrix of the quotient-group K/H.* The order, m, of the latter is relatively prime to the modulus p. Hence by § 1 there exists a matrix μ, of order m and with elements in the field F_p, which transforms E_{11} into its ultimate canonical form C_{11}. Then the compound matrix

$$\begin{pmatrix} \mu & 0 & 0 & \cdots \\ 0 & \mu & 0 & \cdots \\ \cdot & \cdot & \cdot & \cdot \end{pmatrix} \tag{32}$$

transforms the compound matrix (E_{ij}) into $(\mu^{-1}E_{ij}\mu) = (C_{ij})$. Here the C_i have simultaneously their canonical forms:

$$C_{ij} = \begin{pmatrix} c_1^{ij} & 0 & 0 & \cdots \\ 0 & c_2^{ij} & 0 & \cdots \\ \cdot & \cdot & \cdot & \cdot \end{pmatrix}, \tag{33}$$

where c_s^{ij} is a square matrix, whose determinant is irreducible in F_p, and whose elements are linear functions of $\sigma_1^{ij}, \cdots, \sigma_m^{ij}$, with coefficients independent of i, j. Applying to (C_{ij}) a certain transformation which merely permutes the variables, we get

$$\begin{pmatrix} C_1 & 0 & 0 & \cdots \\ 0 & C_2 & 0 & \cdots \\ \cdot & \cdot & \cdot & \cdot \end{pmatrix}, \qquad C_s = (c_s^{ij}) \qquad (i, j = 1, \cdots, n). \tag{34}$$

The problem has therefore been reduced to the normalization of certain matrices C_s ("belonging" to the group G) of order n, the index of H_{p^π} in the largest subgroup (of the given group G) in which H_{p^π} is invariant.

If the Sylow subgroup H_{p^π} is invariant in G, then $n = 1$ and the normalization is complete.

11. From the theorem in § 5 we readily derive an interesting result on group-characters. Let H be a group with the h elements s_i and group-determinant Δ, defined as the determinant of matrix (3). Let G_{p^π} be any Sylow subgroup with the elements g_i and group-determinant D. If we set

$$x_s = 0 \text{ for every element } s \text{ of } H \text{ not in } G_{p^\pi}, \tag{35}$$

Δ becomes equal to an exact power of D. By § 5,

$$D \equiv L^{p^\pi} \pmod{p}, \qquad L = \sum_{i=0}^{p^\pi - 1} x_{g_i}.$$

In our field F_p, let Φ be an irreducible factor of degree f of Δ. Hence, under the assumption (35), Φ becomes a function congruent to L^f modulo p. The coefficient of

$$x_I^{f-1} x_{g_i} \qquad (g_i \neq I) \tag{36}$$

in L^f is f. By definition the coefficient of (36) in Φ is the character $\chi(g_i)$, while $\chi(I) = f$. We may therefore state the

THEOREM. *If χ is the group-character defined by a factor, of degree f and irreducible in the field F_p, of the group-determinant of a finite group H, then*

$$\chi(g) \equiv f \pmod{p} \tag{37}$$

for every element g of period a power of p in H.

The definition of characters is employed also for reducible factors. But

$$\Phi = \Phi' \Phi'' \text{ implies } \chi(g) = \chi'(g) + \chi''(g).$$

Hence (37) *is true also of algebraic group-characters.*

THE UNIVERSITY OF CHICAGO,
May 8, 1907.

EXISTENCE PROOF FOR A FIELD OF EXTREMALS TANGENT TO A GIVEN CURVE*

BY

OSKAR BOLZA

In a recent paper,† Professor BLISS has given sufficient conditions for a minimum of the integral

$$J = \int_{t_0}^{t_1} F(x, y, x', y')\,dt \tag{1}$$

with respect to one-sided variations. His proof is based upon the construction of a *field of extremals tangent to a given curve.* He establishes the existence of such a field first for the special case where all curves considered are representable in the form $y = f(x)$, and then reduces the general case of parameter representation to the former by a point-transformation of the plane.

The object of the following note is to give a *direct proof for the existence of these fields* which play an important part also in other investigations of the calculus of variations.‡

§ 1. *The set of extremals tangent to a given curve.*

The terminology and assumptions concerning the function F being the same as in § 24 of my *Lectures on the Calculus of Variations*, we consider a curve *of class* C'''

$$\mathfrak{C}: \qquad x = \tilde{x}(a), \qquad y = \tilde{y}(a), \qquad A_1 \leqq a \leqq A_2,$$

without multiple points, which lies *in the interior of the region of the* x, y-*plane in which the function* F *is supposed to be of class* C''' for every $(x', y') \neq (0, 0)$, and *satisfies the inequality*

* Presented to the Society March 30, 1907. Received for publication February 27, 1907.

† Transactions of the American Mathematical Society, vol. 5 (1904), p. 477.

‡ Compare LINDEBERG, Mathematische Annalen, vol. 59 (1904), p. 321.

$$F_1[\tilde{x}(a),\ \tilde{y}(a),\ \tilde{x}'(a),\ \tilde{y}'(a)] > 0 \text{ in } (A_1 A_2), \tag{2}$$

where $\tilde{x}' = d\tilde{x}/da$, $\tilde{y}' = d\tilde{y}/da$.

For simplicity, we suppose that the parameter a is the arc of the curve $\tilde{\mathfrak{C}}$ measured from some fixed initial point.

Under these conditions it follows from the general existence theorems* for differential equations applied to the differential equation of the extremals† for the integral (1), that through every point $P(a)$ of the curve $\tilde{\mathfrak{C}}$ one and but one extremal $\mathfrak{E}_a$ can be drawn which is tangent to $\tilde{\mathfrak{C}}$ at P in such a manner that the positive tangents of the two curves coincide. For the parameter t on the extremal $\mathfrak{E}_a$ we may choose the arc of the extremal measured from the point P so that for every value of a the point P corresponds on $\mathfrak{E}_a$ to the value $t = 0$.

If we vary a, we thus obtain a set of extremals

$$x = \phi(t, a), \qquad y = \psi(t, a), \tag{3}$$

for which the functions ϕ, ψ have the following properties:

1) The functions

$$\phi,\ \phi_t,\ \phi_a;\ \psi,\ \psi_t,\ \psi_a$$

are as functions of t and a of class C' in the domain

$$0 \leqq t \leqq l, \qquad A_1 \leqq a \leqq A_2, \tag{4}$$

where l is a sufficiently small positive quantity *independent of* a.‡

2) The functions ϕ, ψ satisfy the following initial conditions:

$$\begin{aligned} \phi(0, a) &= \tilde{x}(a), & \psi(0, a) &= \tilde{y}(a), \\ \phi_t(0, a) &= \tilde{x}'(a), & \psi_t(0, a) &= \tilde{y}'(a). \end{aligned} \tag{5}$$

From (5) we obtain by differentiation

$$\begin{aligned} \phi_a(0, a) &= \tilde{x}'(a), & \psi_a(0, a) &= \tilde{y}'(a), \\ \phi_{ta}(0, a) &= \tilde{x}''(a), & \psi_{ta}(0, a) &= \tilde{y}''(a). \end{aligned} \tag{6}$$

From these equations we derive for the Jacobian

$$\Delta(t, a) = \frac{\partial(\phi, \psi)}{\partial(t, a)}$$

---

* Compare BLISS, Annals of Mathematics, ser. 2, vol. 6 (1905), pp. 49–67.

† Compare KNESER, *Lehrbuch der Variationsrechnung*, §§ 27, 29 and BOLZA, *Lectures on the Calculus of Variations*, § 25, *b*).

‡ Compare the corollary given by BLISS in the article on differential equations just referred to, p. 53, at the end of section 1.

the result:

$$\Delta(0, a) = 0, \qquad \Delta_t(0, a) = \frac{1}{r} - \frac{1}{\tilde{r}}, \tag{7}$$

if we denote by $1/r$ and $1/\tilde{r}$ the curvature at the point P of the curve $\mathfrak{E}_a$ and of the curve $\tilde{\mathfrak{E}}$ respectively.

We make the *further assumption* that

$$\frac{1}{r} - \frac{1}{\tilde{r}} \neq 0$$

along $\tilde{\mathfrak{E}}$, and in order to fix the ideas we suppose * that

$$\frac{1}{r} - \frac{1}{\tilde{r}} > 0. \tag{8}$$

From this additional assumption it follows that two positive quantities $l_0 \leqq l$ and m can be determined so that

$$\Delta(t, a) \geqq tm \tag{9}$$

in the domain

$$0 \leqq t \leqq l_0, \qquad A_1 \leqq a \leqq A_2. \tag{10}$$

For if we define the function $\chi(t, a)$ for the domain (4) by the equations

$$\chi(t, a) = \begin{cases} \dfrac{\Delta(t, a)}{t}, & \text{when } t \neq 0, \\ \Delta_t(0, a), & \text{when } t = 0, \end{cases}$$

it is easily seen that $\chi(t, a)$ is continuous in the domain (4), and since moreover $\chi(0, a) > 0$ in $(A_1 A_2)$, it follows that a positive quantity $l_0 \leqq l$ can be assigned such that $\chi(t, a) > 0$ in the domain (10). If we denote by m the minimum of $\chi(t, a)$ in the domain (10), we obtain (9).

§ 2. *Proof that the set of extremals* (3) *furnishes a field.*

We now choose two quantities a_1, a_2 so that

$$A_1 < a_1 < a_2 < A_2$$

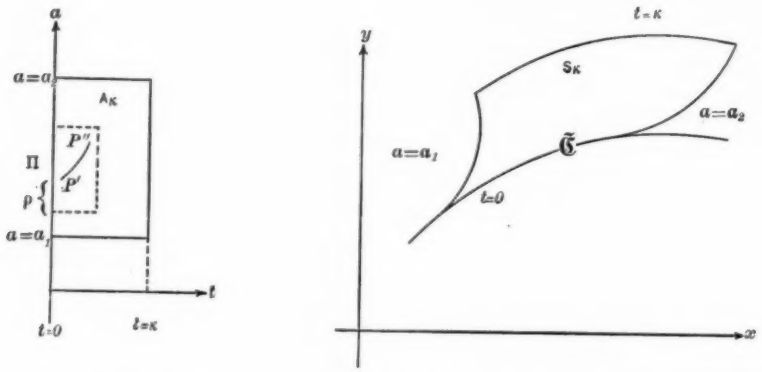
and propose to prove that under the assumptions enumerated in § 1, *a positive quantity* $k \leqq l_0$ *can be assigned such that the equations* (3) *define a one-to-one correspondence between the rectangle*

* In order that the curve $\tilde{\mathfrak{E}}$ may furnish a minimum for the integral (1) with respect to one-sided variations on the left of $\tilde{\mathfrak{E}}$, it is necessary that $1/r - 1/\tilde{r} \geqq 0$; compare BOLZA, *Lectures*, p. 194.

$$\mathbf{A}_\kappa: \qquad 0 \leqq t \leqq \kappa, \qquad a_1 \leqq a \leqq a_2$$

in the t, a-plane and its image $\mathbf{S}_\kappa$ in the x, y-plane.

We suppose that it were not so; that is we suppose that, however small κ may be taken, there always exists in $\mathbf{A}_\kappa$ at least one pair of distinct points whose images in the x, y-plane coincide. Reasoning then exactly as in the proof for the exist-



ence of a field which I have given in § 34 of my *Lectures* for the case where $\Delta(t, a) \neq 0$, we reach the result that under this hypothesis there would exist a point $\Pi(t = 0, a = \alpha)$ in the rectangle $\mathbf{A}_\kappa$ such that every vicinity of Π contains at least one pair of distinct points of $\mathbf{A}_\kappa$ whose images in the x, y-plane coincide.

We are going to prove that this leads to a contradiction with the inequality (9).

For this purpose we notice that our assumptions concerning the curve $\tilde{\mathfrak{C}}$ imply * that $\tilde{x}'(\alpha)$, $\tilde{y}'(\alpha)$ are not both zero; let $\tilde{x}'(\alpha) \neq 0$, or as we may write on account of (6),

$$\phi_a(0, \alpha) \neq 0.$$

We may then apply DINI's theorem on implicit functions to the function $\phi(t, a)$ and the point $t = 0$, $a = \alpha$. From this theorem it follows † that below any

* Compare the definition of "curve of class C''" on p. 116 of my *Lectures*.

† Choose $d > 0$ so that $\phi(t, a)$ is of class C' and $\phi_a(t, a) \neq 0$ for $|t| \leqq d$, $|a - \alpha| \leqq d$. Let A be the maximum of $|\phi_t(t, a)|$, B the minimum of $|\phi_a(t, a)|$ in this domain. Choose $0 < d_0 < d$ and

$$\sigma < d_1 - d_0, \ \rho < d_0, \ \frac{\sigma}{2}, \ \frac{B\sigma}{2A}.$$

Compare PEANO-GENOCCHI, *Differentialrechnung und Grundzüge der Integralrechnung*, pp. 138–141.

preassigned positive quantity δ two positive quantities ρ and σ can be determined having the following properties: If $P'(t', a')$ and $P''(t'', a'')$ be any two distinct points of the vicinity (ρ) of the point $\Pi(0, \alpha)$ for which

$$\phi(t', a') = \phi(t'', a''), \tag{11}$$

then in the first place $t'' \neq t'$ (say $t' < t''$) and in the second place the two points P', P'' can be joined by a curve representable in the form

$$a = \mathfrak{a}(t), \qquad t' \leqq t \leqq t'',$$

such that

$$\phi[t, \mathfrak{a}(t)] = \phi(t', a') \text{ for } t' \leqq t \leqq t''. \tag{12}$$

The function $\mathfrak{a}(t)$ is of C', and satisfies the inequality

$$|\mathfrak{a}(t) - a')| < \sigma \text{ for } t' \leqq t \leqq t''$$

and the initial conditions

$$\mathfrak{a}(t') = a', \qquad \mathfrak{a}(t'') = a''. \tag{13}$$

Differentiating (12) we obtain

$$\phi_t[t, \mathfrak{a}(t)] + \phi_a[t, \mathfrak{a}(t)]\,\mathfrak{a}'(t) = 0. \tag{14}$$

On the other hand, it follows from the characteristic property of the point Π that there exists at least one pair of distinct points P', P'' in the domain

$$0 \leqq t < \rho, \qquad |a - \alpha| < \rho$$

for which not only (11) holds but at the same time

$$\psi(t', a') = \psi(t'', a''). \tag{15}$$

For such a pair of points the function $\psi[t, \mathfrak{a}(t)]$ is of class C' in $(t't'')$ and takes, according to (13) and (15), the same value for $t = t'$ and $t = t''$. Hence its derivative must vanish at least for one value $t = \tau$ between t' and t'':

$$\psi_t(\tau, \mathfrak{a}(\tau)] + \psi_a[\tau, \mathfrak{a}(\tau)]\,\mathfrak{a}'(\tau) = 0.$$

Combining this equation with the equation derived from (14) by putting $t = \tau$, we obtain the result:

$$\Delta[\tau, \mathfrak{a}(\tau)] = 0.$$

But if we take ρ and σ sufficiently small, the point $t = \tau$, $a = \mathfrak{a}(\tau)$ lies in the domain (10); moreover, τ is positive since $0 \leqq t' < \tau < t''$. Hence we have indeed reached a contradiction with the inequality (9), and therefore the statement enunciated at the beginning of this section is proved.

The image $\mathfrak{L}$ of the boundary of the rectangle $\mathbf{A}_\varkappa$ is a continuous closed curve without multiple points. According to a theorem due to SCHÖNFLIES,* the point-set $\mathbf{S}_\varkappa$ is therefore identical with the interior of $\mathfrak{L}$ together with the curve $\mathfrak{L}$ itself.

THE UNIVERSITY OF CHICAGO,
February 27, 1907.

*Göttinger Nachrichten, 1899, p. 282; compare also OSGOOD, ibid., 1900, p. 94; and BERNSTEIN, ibid., 1900, p. 98.

A NEW FORM OF THE SIMPLEST PROBLEM OF THE CALCULUS OF VARIATIONS*

BY

GILBERT AMES BLISS

The simplest problem of the calculus of variations in the plane has been studied in two different forms. In the earlier one the curves considered are represented by equations

$$y = y(x),$$

where $y(x)$ is a single-valued function, and the integral to be minimized is

$$I = \int f(x, y, y')\,dx. \tag{1}$$

In the later form, which seems first to have been studied in detail by WEIERSTRASS, the curves are taken in parametric representation

$$C: \qquad x = \phi(t), \qquad y = \psi(t), \tag{2}$$

and the integral in the form

$$I = \int F(x, y, x_t, y_t)\,dt, \tag{3}$$

where the function F is positively homogeneous in the derivatives x_t and y_t. Both methods are open to objections, the former because it is inconvenient to apply to curves for which the function $y(x)$ is not single valued or for which the slope becomes infinite. In fact, no complete discussion has been attempted for the problem of minimizing the integral (1) without restricting the functions $y(x)$ to be single-valued and to have a finite derivative. The second form of the problem is complicated by the fact that in all the calculations involved the condition of homogeneity

$$F(x, y, \kappa x_t, \kappa y_t) = \kappa F(x, y, x_t, y_t) \qquad (\kappa > 0),$$

must be kept in mind and frequently applied.

In a paper entitled *A generalization of the notion of angle* † the writer has introduced a third form of the problem, in which the integral involves only invariants of the curve (2) under change of parameter-representation. The integral there given is

* Presented to the Society April 27, 1907. Received for publication April 8, 1907.

† Transactions of the American Mathematical Society, vol. 7 (1906), p. 184.

$$I = \int f(x, y, \tau)\, ds, \tag{4}$$

where s is the length of arc along the curve (2), and τ is the angle made by the curve with the x-axis, which is defined by the equations

$$\cos\tau = \frac{x_t}{\sqrt{x_t^2 + y_t^2}}, \qquad \sin\tau = \frac{y_t}{\sqrt{x_t^2 + y_t^2}}. \tag{5}$$

The integral (3) may be easily put into the form (4), for on account of the homogeneity of F,

$$F(x, y, x_t, y_t) = F(x, y, \cos\tau, \sin\tau)\sqrt{x_t^2 + y_t^2}.$$

But in (4) the function $f(x, y, \tau)$ need not be periodic as is the case with the right member of the last equation. Besides this generalization, the advantages of having the integral in the form (4) are that one can use the parametric representation for curves and at the same time avoid the troublesome homogeneity condition connected with the Weierstrassian form of the problem. It must be admitted that the derivation of the important formulas is not in general as simple as is the case for the integral (1), but otherwise the advantages of both of the older methods seem to be retained. The new form of the problem is especially advantageous in connection with the inverse problem of the calculus of variations and its application to geometry as made by HAMEL,* and in determining the invariants of the problem under point transformations.

In the paper by the writer referred to above only the Euler equation and the equations defining transversality, which were necessary in finding the generalization of angle, were developed. It is proposed here to complete the theory of the problem connected with the integral (4) by deriving the forms of the first and second variations and the functions which are commonly used in proving the necessary conditions in the calculus of variations. The derivation of the conditions themselves can be made as is usual in the books on the subject.†

§ 1. *The first variation.*

Suppose then that the integral to be studied is

$$I = \int f(x, y, \tau)\sqrt{x_t^2 + y_t^2}\, dt,$$

where

$$\sqrt{x_t^2 + y_t^2} = \frac{ds}{dt}$$

* *Über die Geometrieen in denen die Geraden die Kürzesten sind*, Mathematische Annalen, vol. 57 (1903), p. 231.

† See especially BOLZA, *Lectures on the Calculus of Variations.*

and where the function $f(x, y, \tau)$ is of class C^{iv}* in a region R of (x, y, τ)-points having an interior. The curves considered will all be represented in the parametric representation (2) where ϕ and ψ are functions of class C^{iv} and define values (x, y, τ) lying interior to the region R for all values of t in the interval $t_0 \leqq t \leqq t_1$. As the value of τ is to be well-defined for every value of t, it is evident that the expression $x_t^2 + y_t^2$ must be everywhere different from zero, and the curve will therefore be supposed to have no singular points.

If the integral I is taken along one of the family of curves

$$C_a: \qquad x = \phi(t, a), \qquad y = \psi(t, a) \tag{6}$$

where $\phi(t, a)$ and $\psi(t, a)$ are of class C^{iv} in the region

$$t_0 \leqq t \leqq t_1, \qquad |a| \leqq k \qquad (k > 0),$$

and for $a = 0$ define the curve (2), then the value of I becomes a function $I(a)$ which for $a = 0$ reduces to the value I taken along C. The first derivative is

$$\frac{dI}{da} = \int_{t_0}^{t_1} \left\{ f_x x_a + f_y y_a + f_\tau \tau_a + f \frac{x_t x_{at} + y_t y_{at}}{x_t^2 + y_t^2} \right\} \sqrt{x_t^2 + y_t^2}\, dt.$$

In this and following formulas literal subscripts denote partial differentiation, except the subscript s which means that there has first been a partial differentiation with respect to t and then a division by $s_t = \sqrt{x_t^2 + y_t^2}$.

After substitution of the value of τ_a calculated from equations (5) and (6), the last equation becomes

$$\frac{dI}{da} = \int_{s_0}^{s_1} \{ f_x x_a + f_y y_a + (f \cos\tau - f_\tau \sin\tau) x_{as} + (f \sin\tau + f_\tau \cos\tau) y_{as} \}\, ds,$$

or by the usual partial integration of the calculus of variations,

$$\frac{dI}{da} = \int_{s_0}^{s_1} \{ P x_a + Q y_a \}\, ds + [(f \cos\tau - f_\tau \sin\tau) x_a + (f \sin\tau + f_\tau \cos\tau) y_a]_{s_0}^{s_1},$$

where

$$P = f_x - \frac{d}{ds}(f \cos\tau - f_\tau \sin\tau), \qquad Q = f_y - \frac{d}{ds}(f \sin\tau + f_\tau \cos\tau).$$

By an easy calculation P and Q are found to satisfy the relation

$$P \cos\tau + Q \sin\tau = 0,$$

so that a new expression T can be introduced by means of the equations

$$P = T \sin\tau, \qquad Q = -T \cos\tau.$$

* A function of class $C^{(n)}$ is by definition one which is itself continuous and has continuous derivatives up to and including those of the nth order. The assumptions on $f(x, y, \tau)$ and the curve (2) insure the existence of all the derivatives used in the following developments.

The final form of the first derivative is then

$$(7)\quad \frac{dI}{d\alpha}=\int_{s_0}^{s_1} wT\,ds+[(f\cos\tau-f_\tau\sin\tau)x_\alpha+(f\sin\tau+f_\tau\cos\tau)y_\alpha]_{s_1}^{s_0},$$

where

$$(8)\qquad w=x_\alpha\sin\tau-y_\alpha\cos\tau,$$

$$(9)\quad T(x,y,\tau,\tau_s)=f_x\sin\tau-f_y\cos\tau+f_{x\tau}\cos\tau+f_{y\tau}\sin\tau+(f+f_{\tau\tau})\tau_s.$$

In case the curves (6) all pass through the two end-points 0 and 1 of the arc C, one sees readily by putting $t=t_0$ or $t=t_1$ in equations (6) and differentiating for α, that x_α, y_α both vanish at the points 0 and 1, so that *for a set of curves* (6) *all passing through two fixed points* 0 *and* 1 *the derivative* $dI/d\alpha$ *has the value*

$$\frac{dI}{d\alpha}=\int_{s_0}^{s_1} wT\,ds,$$

where w *and* T *are the expressions in* (8) *and* (9) *above.*

The function w depends upon the family of curves (6). But if w is given in advance, a family can always be found including the arc C for $\alpha=0$ and having the given value w. It is only necessary to take curves in the form

$$(10)\qquad x=\phi(t)+\alpha\xi,\qquad y=\psi(t)+a\eta,$$

where ξ and η are to be determined by the equations

$$\xi\sin\tau-\eta\cos\tau=w,$$

$$\xi\cos\tau+\eta\sin\tau=0.$$

In particular, if w vanishes for $t=t_0$ and $t=t_1$, the curves (10) will all pass through two fixed points.

By methods well-known in the calculus of variations it follows then that *any curve which joins two given fixed points and minimizes the integral* I *must be a solution of Euler's differential equation*

$$(11)\qquad T(x,y,\tau,\tau_s)=0.$$

§ 2. *The second variation.*

For the case when the end-points are fixed the second derivative $d^2I/d\alpha^2$ can be calculated from the expression

$$\frac{dI}{d\alpha}=\int_{t_0}^{t_1} T\cdot w\sqrt{x_t^2+y_t^2}\,dt.$$

It is seen readily that the derivative of the integrand is

$$(12)\quad w\{T_x x_\alpha+T_{\dot x}y_\alpha+T_\tau\tau_\alpha+(f+f_{\tau\tau})\tau_{s\alpha}\}\sqrt{x_t^2+y_t^2}+T\frac{\partial}{\partial\alpha}(w\sqrt{x_t^2+y_t^2}).$$

The value of this second derivative is desired for $\alpha = 0$, that is, along the curve C. If C satisfies the first necessary condition for a minimum, $T = 0$, the term with the factor T disappears. Furthermore from (9) and (5)

$$(13) \qquad T_\tau = f_s + f_{\tau\tau s}, \qquad \tau_a = (x_a \cos\tau + y_a \sin\tau)\tau_s - w_s,$$

$$\tau_{sa} = \frac{\partial}{\partial\alpha}\frac{\tau_t}{\sqrt{x_t^2 + y_t^2}} = (x_a \cos\tau + y_a \sin\tau)\tau_{ss} - w_{ss} - \tau_s^2 w,$$

so that for $\alpha = 0$ the expression (12) becomes

$$(14) \quad w\left\{\left(T_x + \cos\tau\frac{d}{ds}f_1\tau_s\right)x_a + \left(T_y + \sin\tau\frac{d}{ds}f_1\tau_s\right)y_a - f_1\tau_s^2 w - \frac{d}{ds}f_1 w_s\right\}$$

where

$$f_1 = f + f_{\tau\tau}.$$

But along the curve C

$$\left(T_x + \cos\tau\frac{d}{ds}f_1\tau_s\right)\cos\tau + \left(T_y + \sin\tau\frac{d}{ds}f_1\tau_s\right)\sin\tau = T_s = 0.$$

When this equation and

$$\left(T_x + \cos\tau\frac{d}{ds}f_1\tau_s\right)\sin\tau - \left(T_y + \sin\tau\frac{d}{ds}f_1\tau_s\right)\cos\tau = T_x\sin\tau - T_y\cos\tau$$

are solved for the parentheses and the results substituted in (14), one finds without difficulty

$$(15) \qquad \left.\frac{d^2 I}{d\alpha^2}\right|^{\alpha=0} = \int_{s_0}^{s_1} w\left(f_2 w - \frac{d}{ds}f_1 w_s\right)ds = \int_{s_0}^{s_1}\{f_1 w_s^2 + f_2 w^2\}\,ds,$$

where

$$f_2(x, y, \tau, \tau_s) = T_x\sin\tau - T_y\cos\tau - f_1\tau_s^2.$$

The latter part of (15) is derived from the former by a simple partial integration, the terms outside the integral disappearing because for fixed end-points w vanishes at $s = s_0$ and $s = s_1$. These, however, are the well-known forms for the second variation, a factor $\alpha^2/2$ only being omitted, from which one derives the two following conditions.

LEGENDRE'S NECESSARY CONDITION. *The function $f_1 = f + f_{\tau\tau}$ must be $\geqq 0$ for a minimum, or $\leqq 0$ for a maximum, at every point of a minimizing arc C.*

JACOBI'S NECESSARY CONDITION. *If the arc C satisfies Euler's equation and if furthermore $f_1 \neq 0$ along C, then a further necessary condition for a minimum or maximum is that no integral w of Jacobi's equation,*

$$f_2 w - \frac{d}{ds}f_1 w_s = 0,$$

exists which vanishes more than once in the interval $s_0 \leqq s < s_1$, the arguments of $f_2(x, y, \tau, \tau_s)$ and $f_1(x, y, \tau)$ in this equation being those corresponding to the arc C.

§ 3. *Hilbert's invariant integral and the E-function of Weierstrass.*

As soon as it is known that the arc C is a solution of Euler's equation along which $f_1 \neq 0$ and which satisfies Jacobi's condition, the existence theorems for differential equations enable one to imbed C in a field of solutions of Euler's equation. In fact, if a point 2 is taken near enough to 0 on C and having the order 201 with the points 0 and 1, then the solutions of (11) which pass through the point 2 will be a one-parameter family of curves which simply cover a portion of the plane in the neighborhood of the arc C.* Their equations can be found in the form

$$x = \phi(t, \alpha), \qquad y = \psi(t, \alpha) \qquad (0 \leqq t \leqq T,\ |\alpha| \leqq k), \tag{16}$$

where the value $t = 0$ corresponds to the point 2 on each curve, and where for $\alpha = 0$ the arc C itself is represented. Under the assumptions originally made upon the function $f(x, y, \tau)$ the functions ϕ and ψ will be of class C^{iv}.

In the field surrounding C the equations (16) are uniquely solvable for t and α in terms of x and y,

$$t = t(x, y), \qquad \alpha = \alpha(x, y), \tag{17}$$

and by implicit function theory these functions will be of class C' at least.

HILBERT's invariant integral may be derived from the integral

$$I = \int_0^t f(x, y, \tau)\, ds$$

taken along any one of the solutions (16) from the point $2\,(t = 0)$ to a point corresponding to the value t. This integral is a function of t and α, or by means of equations (17), a function of x and y in the field. If one takes the integral

$$I^* = \int_{u_3}^{u_4} \left\{ \frac{\partial I}{\partial x}\, x_u + \frac{\partial I}{\partial y}\, y_u \right\} du \tag{18}$$

over any path

$$\bar{C}: \qquad x = x(u), \qquad y = y(u), \tag{19}$$

in the field joining two given points 3, 4, it is evident that its value will always be the same however the path is chosen. The integral (18) is indeed HILBERT's invariant integral. It is desirable, however, to find more definitely the expressions for $\partial I/\partial x$ and $\partial I/\partial y$. By steps like those of § 1 it is seen that

$$\frac{\partial I}{\partial t} = f(x, y, \tau)\sqrt{x_t^2 + y_t^2},$$

$$\frac{\partial I}{\partial \alpha} = (f\cos\tau - f_\tau \sin\tau)x_\alpha + (f\sin\tau + f_\tau\cos\tau)y_\alpha.$$

* In the case when the function $f(x, y, \tau)$ is periodic with the period 2π in τ the family of solutions through the point 0 has the same property. See BLISS, *The construction of a field of extremals about a given point*, Bulletin of the American Mathematical Society, vol. 13 (1907), p. 321.

Since the curves (16) are all solutions of Euler's equation the expression T vanishes identically, while x_α and y_α vanish at the point 0, so that in equation (17) only the terms outside the integral sign at the point t are left. It follows therefore that

$$\frac{\partial I}{\partial x} = f(x, y, \tau)\sqrt{x_t^2 + y_t^2}\frac{\partial t}{\partial x} + [(f\cos\tau - f_\tau \sin\tau)x_\alpha + (f\sin\tau + f_\tau\cos\tau)y_\alpha]\frac{\partial \alpha}{\partial x},$$

with a similar equation for the y-derivative. The integrand of HILBERT's invariant integral (18) is evidently

$$f(x, y, \tau)\sqrt{x_t^2 + y_t^2}\frac{dt}{du} + [(f\cos\tau - f_\tau\sin\tau)x_\alpha + (f\sin\tau + f_\tau\cos\tau)y_\alpha]\frac{d\alpha}{du},$$

which can be easily changed into the form

$$(f\cos\tau - f_\tau\sin\tau)x_u + (f\sin\tau + f_\tau\cos\tau)y_u$$

since

$$\sqrt{x_t^2 + y_t^2} = x_t\cos\tau + y_t\sin\tau,$$

$$x_t\frac{dt}{du} + x_\alpha\frac{d\alpha}{du} = x_u, \qquad y_t\frac{dt}{du} + y_\alpha\frac{d\alpha}{du} = y_u.$$

If an angle σ is defined by the equations

$$\cos\sigma = \frac{x_u}{\sqrt{x_u^2 + y_u^2}}, \qquad \sin\sigma = \frac{y_u}{\sqrt{x_u^2 + y_u^2}},$$

Hilbert's integral may be written in the form

$$I^* = \int \{f\cos(\sigma - \tau) + f_\tau\sin(\sigma - \tau)\}\sqrt{x_u^2 + y_u^2}\,du,$$

where σ and τ are the direction-angles of the arbitrarily chosen curve (19) *and one of the solutions* (16) *of Euler's equation, respectively, at their point of intersection in the field. This integral has the fundamental properties that its value taken between any two points of the field is independent of the path over which it is taken, and on one of the solutions* (16) *of Euler's equation it has the same value as the original integral I.*

The latter property follows at once, since along an extremal $\sigma = \tau$.

The WEIERSTRASS E-function is the function

$$E(x, y, \sigma, \tau) = f\cos(\sigma - \tau) + f_\tau\sin(\sigma - \tau) - f(x, y, \sigma)$$

occurring in the integrand of the difference

$$I(\bar{C}) - I^*(\bar{C}) = \int E\sqrt{x_u^2 + y_u^2}\,du, \tag{20}$$

taken along any curve $\bar{C}$ of the field. If the ends of the arc over which the integration is taken lie on the arc C, this difference has the value

$$I(\bar{C}) - I^*(\bar{C}) = I(\bar{C}) - I(C) \tag{21}$$

on account of the properties of I^* above mentioned.

By means of special variations WEIERSTRASS derived the following fourth necessary condition involving the E-function.*

WEIERSTRASS'S NECESSARY CONDITION. *If the arc C is a solution of Euler's equation which minimizes the integral I with respect to any curve lying in a properly chosen neighborhood of C and joining its end-points, then the inequality*

$$E(x, y, \sigma, \tau) \geqq 0$$

must hold for every set of values (x, y, τ) corresponding to a point on C and for all values of σ. In case there is a maximum E must be $\leqq 0$.

It is well known that the four necessary conditions given above do not at the same time insure a minimum or a maximum. By strengthening the inequalities involved in them, however, new sets of conditions may be derived in various ways which with the help of a field and the equations (20), (21) can be proved sufficient. For the sake of completeness one of the theorems which define sufficient conditions is given here.

If the arc C is a solution of Euler's equation along which Legendre's and Weierstrass's conditions hold in the stronger forms $f_1 > 0$ and $E > 0$, and if furthermore Jacobi's condition is satisfied not only in the interval $s_0 \leqq s < s_1$ but also in $s_0 \leqq s \leqq s_1$, then a neighborhood of C can always be found in which this arc actually minimizes the integral I.

§ 4. *Variable endpoints.*

In the preceding sections the formulæ derived were for the case of fixed end-points. When a minimizing arc C is sought, however, which joins a fixed point 0 with a fixed curve D, a new condition on the directions of C and D at their point of intersection must be satisfied and the JACOBI condition takes on a somewhat different form. For the purpose of deriving these conditions let the representation of the family of curves (6) be chosen, as is usually done, so that their intersections with the curve D all occur at the fixed parameter-value t_1. Then the curve D is represented by the equations

$$x = \phi(t_1, \alpha), \qquad y = \psi(t_1, \alpha),$$

in which α is to be regarded as the parameter. The integral I, taken over a curve of this set and regarded as a function of α, has the first derivative

* See BOLZA, *Lectures on the calculus of variations*, p. 138.

$$(22)\qquad \frac{dI}{d\alpha} = \int_{s_0}^{s_1} wT ds + \left[\sqrt{x_\alpha^2 + y_\alpha^2}\,\{f\cos(\sigma-\tau) + f_\tau \sin(\sigma-\tau)\}\right]^{t=t_1},$$

a formula derived from equation (7), in which σ denotes the direction of the curve D derived from the equations

$$\cos\sigma = \frac{x_\alpha}{\sqrt{x_\alpha^2 + y_\alpha^2}}, \qquad \sin\sigma = \frac{y_\alpha}{\sqrt{x_\alpha^2 + y_\alpha^2}},$$

for $t = t_1$. The terms outside the integral in (7) vanish at the point 0 exactly as before, but not necessarily at $t = t_1$. If the arc C is to minimize I with respect to all curves joining the point 0 with the curve D, it must certainly minimize I with respect to curves which join 0 with the intersection of C and D, and consequently must be a solution of Euler's equation as in § 1. In that case the integral part of the expression (22) vanishes for $\alpha = 0$. But the rest must also vanish for $\alpha = 0$, and so the following condition must be satisfied.

CONDITION OF TRANSVERSALITY. *If the arc C is a solution of Euler's equation which minimizes the integral I with respect to curves joining a fixed point 0 with a fixed curve D, then at its intersection with D the equation*

$$(23)\qquad f\cos(\sigma-\tau) + f_\tau \sin(\sigma-\tau) = 0$$

*must also be satisfied, where σ denotes the direction of D and τ the direction of C.**

For the second derivative $d^2I/d\alpha^2$, the integral part has the form given first in equation (15), and it remains only to find the derivative of the terms in (22) which are outside the integral sign. For $\alpha = 0$ the expression

$$\left[f\cos(\sigma-\tau) + f_\tau \sin(\sigma-\tau)\right]\frac{\partial}{\partial\alpha}\sqrt{x_\alpha^2 + y_\alpha^2}$$

vanishes on account of equation (23). The rest of the derivative, omitting a factor $\sqrt{x_\alpha^2 + y_\alpha^2}$, is

$$(24)\qquad \begin{aligned}&\left[-f\sin(\sigma-\tau) + f_\tau\cos(\sigma-\tau)\right](\sigma_\alpha - \tau_\alpha)\\ &+(f_x x_\alpha + f_y y_\alpha + f_\tau \tau_\alpha)\cos(\sigma-\tau) + (f_{\tau x} x_\alpha + f_{\tau y} y_\alpha + f_{\tau\tau}\tau_\alpha)\sin(\sigma-\tau).\end{aligned}$$

In this expression, from (13) and the theory of plane curves,

$$(25)\qquad \tau_\alpha = \sqrt{x_\alpha^2 + y_\alpha^2}\cos(\sigma-\tau)\cdot\tau_s - w_s, \qquad \sigma_\alpha = \frac{\sqrt{x_\alpha^2 + y_\alpha^2}}{r},$$

* The equation (23) is satisfied if

$$\cos\sigma = \frac{f\sin\tau + f_\tau\cos\tau}{\sqrt{f^2 + f_\tau^2}}, \qquad \sin\sigma = \frac{-f\cos\tau + f_\tau\sin\tau}{\sqrt{f^2 + f_\tau^2}},$$

as in the writer's earlier paper referred to above.

where r is the radius of curvature of the curve D. With the help of (24) and (25) the value of $d^2I/d\alpha^2$ is found to be

$$(26)\quad \frac{d^2I}{d\alpha^2} = \int_{s_0}^{s_1} w\left(f_2 w - \frac{d}{ds} f_1 w_s\right) ds + \left[(x_a^2 + y_a^2)\left(\frac{P_1}{r} + P_2 + Q\frac{w_s}{w}\right)\right]^{t=t_1},$$

where

$$P_1 = -f\sin(\sigma-\tau) + f_\tau \cos(\sigma-\tau),$$

$$P_2 = (f_x\cos\sigma + f_y\sin\sigma)\cos(\sigma-\tau) + (f_{\tau x}\cos\sigma + f_{\tau y}\sin\sigma)\sin(\sigma-\tau) + f_1\sin(\sigma-\tau)\cos(\sigma-\tau)\tau_s,$$

$$Q = -f_1\sin^2(\sigma-\tau).$$

But by means of the expression (26), the Jacobi condition can be proved * in the following form.

JACOBI'S NECESSARY CONDITION. *If the arc C is a solution of Euler's equation joining the fixed point 0 with the fixed curve D, and along which $f_1 \neq 0$, then a further necessary condition for a minimum or a maximum is that no solution w of Jacobi's equation*

$$f_2 w - \frac{d}{ds} f_1 w_s = 0$$

exists which satisfies the condition

$$\frac{P_1}{r} + P_2 + Q\frac{w_2}{w} = 0$$

at the intersection of C and D and vanishes in the interval $s_0 < s \leqq s_1$.

The alterations which must be made in the necessary conditions for the case of variable end-points in order to get sufficient conditions are so similar to the ones given in § 3 that the corresponding theorem need not be stated here.

PRINCETON, *March*, 1907.

* See BLISS, *The second variation of a definite integral when one end-point is variable*, Transactions of the American Mathematical Society, vol. 3 (1902), p. 132.

ON CERTAIN ISOTHERMIC SURFACES*

BY

ARCHER EVERETT YOUNG

I. *Introduction.*

§ 1. *Isothermic Surfaces.* CHRISTOFFEL proposed the problem of determining all pairs of surfaces, S and S_1, for which a one-to-one correspondence of their points can be established, such that the tangent planes at corresponding points will be parallel and the angles between corresponding lines will be equal. His discussion† of the problem shows that, aside from minimal and certain imaginary surfaces, the only solutions are surfaces which, when referred to lines of curvature, have as the expressions for their linear elements, respectively,

$$ds^2 = \lambda(du^2 + dv^2), \qquad ds_1^2 = \frac{1}{\lambda}(du^2 + dv^2),$$

where λ is a function of u and v. CHRISTOFFEL's problem, then, reduces to that of finding all possible isothermic surfaces.

No great progress has been made in the investigation of these surfaces, although WEINGARTEN‡ and DARBOUX§ have shown that they can be defined by a non-linear partial differential equation of the fourth order. The only solutions carried to completion have been obtained by imposing one or more additional conditions; the resulting surfaces have thus been very particular solutions of the general problem.‖

§ 2. *Problem discussed in this paper.* The characteristic form for the linear element of many of the isothermic surfaces, when referred to lines of curvature, is well known.

Those of the quadrics and BONNET surfaces¶ are respectively

$$ds^2 = (u+v)\left(\frac{du^2}{U} + \frac{dv^2}{V}\right), \qquad ds'^2 = \frac{1}{(u+v)^2}\left(\frac{du^2}{U_1} + \frac{dv^2}{V_1}\right),$$

* Presented to the Society (Chicago), December 29, 1905. Received for publication, August 1, 1906.

† Crelle, vol. 57 (1867), pp. 218–228.

‡ Sitzungsberichte der Akademie der Wissenschaften zu Berlin, vol. 2 (1883), p. 1163.

§ *Théorie des surfaces*, vol. 2, p. 248.

‖ DARBOUX, loc. cit., vol. 2, p. 246.

¶ BONNET, Journal de L'Ecole Polytechnique, vol. 42, pp. 121 and 132–151.

where U and U_1 are certain functions of u, and V and V_1 of v. It is clear from CHRISTOFFEL's results that there must be other surfaces corresponding respectively to the expressions

$$ds_1^2 = \frac{1}{(u+v)}\left(\frac{du^2}{U} + \frac{dv^2}{V}\right), \qquad ds'^2_1 = (u+v)^2\left(\frac{du^2}{U_1} + \frac{dv^2}{V_1}\right).$$

Evidently the four preceding expressions are included under the general form

$$ds^2 = (u+v)^K\left(\frac{du^2}{U} + \frac{dv^2}{V}\right), \tag{1}$$

where K is an arbitrary constant, and U and V arbitrary functions of u and v, respectively.

We have undertaken in this paper to determine the surfaces whose linear elements can be written in this form, the lines of reference being lines of curvature. The remainder of this first section is given up to the analytical discussion of the equations involved. It appears that the only surfaces, with the exception of the plane and cylindrical surfaces, that satisfy the conditions of the problem, are those associated with the BONNET surface, the quadrics, or the sphere referred to confocal spherical ellipses.

The associates of the BONNET surfaces are discussed in section II, and the others in section III of this paper. In the latter section, also, attention is called to an apparent oversight of BONNET, occurring in his discussion of the characteristic form for the linear element of the quadrics.

§ 3. *Problem reduced to the solution of a single equation.* The problem of determining all surfaces having their linear element in the form (1), when the lines of reference are lines of curvature, reduces to that of finding all solutions of the following system of equations: *

$$\frac{\partial M}{\partial v} - \frac{\partial N}{\partial u} = PQ, \qquad \frac{\partial P}{\partial v} + MQ = 0, \qquad \frac{\partial Q}{\partial u} - NP = 0, \tag{2}$$

where P and Q are functions of u and v to be determined, and M and N, for the linear element (1), have the values:

$$M = -\frac{K\sqrt{\frac{V}{U}}}{2(u+v)}, \qquad N = \frac{K\sqrt{\frac{U}{V}}}{2(u+v)}.$$

From (2) it follows easily that

$$\frac{\partial(P^2+M^2)}{\partial v} = 2M\frac{\partial N}{\partial u}, \qquad \frac{\partial(Q^2+N^2)}{\partial u} = 2N\frac{\partial M}{\partial v}.$$

* BONNET, loc. cit., p. 133.

Substituting these values of M and N and integrating, we have

$$P^2=\frac{K^2}{4}\left[\frac{U'}{(u+v)}-\frac{U+V}{(u+v)^2}+U_1\right],\quad Q^2=\frac{K^2}{4}\left[\frac{V'}{u+v}-\frac{(U+V)}{(u+v)^2}+V_1\right],$$

where the primes denote derivatives, and U_1 and V_1 are new functions of u and v respectively.

The first equation of (2) with these values substituted becomes

$$K^2\left[U'-\frac{U+V}{u+v}+(u+v)U_1\right]\left[V'-\frac{U+V}{u+v}+(u+v)V_1\right]=\left[U'+V'-2\frac{U+V}{u+v}\right]^2, \tag{3}$$

and the system (2) is solved as soon as all possible sets of values of K, U, U_1, V, V_1, which satisfy this equation, are found.

§ 4. *Discussion of equation* (3)*. In discussing equation (3) we shall assume that for some value, α, all the functions $U(\alpha)$, $U_1(\alpha)$, $V(-\alpha)$, and $V_1(-\alpha)$ are analytic.†

Multiplying both sides of (3) by $(u+v)^2$, and then replacing u by α and v by $-\alpha$, we have

$$(K^2-4)\{U(\alpha)+V(-\alpha)\}=0.$$

Hence

$$1^\circ\quad K=\pm 2,\qquad \text{or}\qquad 2^\circ\quad U(\alpha)=-V(-\alpha).$$

Case 1°, when $K=-2$, has been discussed by BONNET.‡ The functions, as determined by him, which satisfy (3) when K has this value, will satisfy (3) evidently when K has the other value. We have § then, for $K=\pm 2$,

$$U=\alpha u^2+\alpha' u+\alpha'',\qquad U_1=-\tfrac{1}{2}U'',$$

$$V=\text{arbitrary function of } v,\qquad V_1=\text{function of } v \text{ and } V.$$

Case 2° needs further consideration. Writing $U=f(u)$, we have

$$U(\alpha)=f(\alpha),\qquad V(-\alpha)=-f(-\alpha),$$

$$U'(\alpha)=f'(\alpha),\qquad V'(-\alpha)=f'(-\alpha),$$

$$U''(\alpha)=f''(\alpha),\qquad V''(-\alpha)=-f''(-\alpha).$$

* The following discussion of equation (3) was suggested by Professor J. E. WRIGHT.

† The question as to whether this assumption limits our results we shall not attempt to settle in this paper. The same point arises in connection with several other equations which the writer is studying and may be discussed by itself in another article.

‡ Loc. cit., p. 133–151.

§ This solution is not based upon the assumption at the beginning of this article with regard to the functions, but is the most general solution of (3) when $K^2=4$.

If u be replaced by α and v by $(-\alpha+h)$, then (3) may be written

$$K^2[\tfrac{1}{2}f''h+U_1h-\tfrac{1}{6}f''''h^2+\cdots][-\tfrac{1}{2}f''h+\tfrac{1}{2}f''''h^2+\cdots \\ -\tfrac{1}{6}f''''h^2+\cdots+hV_1]=[\tfrac{1}{2}f''''h^2-\tfrac{1}{3}f''''h^2+\cdots]^2, \tag{4}$$

where α is the argument of f, and the terms omitted contain cubes and higher powers of h.

Equating coefficients of h^2, we find

$$[\tfrac{1}{2}f''(\alpha)+U_1(\alpha)][-\tfrac{1}{2}f''(\alpha)+V_1(-\alpha)]=0.$$

Assuming, then, that $U_1(\alpha)=-\tfrac{1}{2}f''(\alpha)$, and equating coefficients of h^3, we have

$$f''''(\alpha)[-\tfrac{1}{2}f''(\alpha)+V_1(-\alpha)]=0,$$

so that if $f''''(\alpha)\neq 0$, then $V_1(-\alpha)=\tfrac{1}{2}f''(\alpha)$, which is the condition we passed over in the preceding equation. Finally, the coefficients of h^4 in (4) give

$$(K^2-1)f''''^2(\alpha)=0.$$

Hence two cases are left for consideration:

(i) $$f''''(\alpha)=0,$$

and therefore

$$U=\alpha u^2+\alpha' u+\alpha'', \qquad V=-\alpha v^2+\alpha' v-\alpha'',$$

(ii) $$K=\pm 1, \qquad U_1=-\tfrac{1}{2}U'', \qquad V_1=-\tfrac{1}{2}V''.$$

Case (i) will be considered later.

Case (ii). In order to determine the forms of the functions U and V we shall replace U_1 and V_1 by the values given in (ii), and then write (3) in the form

$$[(U+V)-U'(u+v)+\tfrac{1}{2}U''(u+v)^2][(U+V)-V'(u+v)+\tfrac{1}{2}V''(u+v)^2] \\ =[(U'+V')(u+v)-2(U+V)]^2. \tag{5}$$

Giving v some constant value, for which V is analytic, and putting

$$(U+V)-V'(u+v)+\tfrac{1}{2}V''(u+v)^2=F,$$

we may write (5) in the form

$$F[F-(u+v)F'+\tfrac{1}{2}F''(u+v)^2]=[F'(u+v)-2F]^2;$$

which, by the substitution

$$F=\phi\cdot(u+v)^2,$$

becomes

$$\tfrac{1}{2}\phi\phi''+\frac{\phi\phi'}{(u+v)}=\phi'^2,$$

where the primes denote derivatives with regard to u. Integrating, we find

$$\frac{1}{\phi} = \frac{\psi(v)}{(u+v)} + \chi(v),$$

where ψ and χ are arbitrary functions of v, and hence

$$F = (U+V) - V'(u+v) + \tfrac{1}{2}V''(u+v)^2 = \frac{(u+v)^3}{Au+B},$$

where A and B are functions of v. Finally, solving for U, we have

$$U = a + bu + cu^2 + \frac{(u+d)^3}{hu+k},$$

where a, b, c, etc., are constants, as U must be independent of v. This result is based on the assumption that the function ψ does not vanish.

If $\psi = 0$, then $f''' = 0$, a case already considered. If $\psi \neq 0$, the constants h and k cannot both be zero at the same time.

If $h \neq 0$, with a new set of variables $(u - k/h)$ and $(v + k/h)$, the new function can be written in the form

$$U = \alpha + \alpha' u + \alpha'' u^2 + \frac{\alpha'''}{u}.$$

If $h = 0$ and $k \neq 0$, the expression for the function may be written

$$U = \alpha + \alpha' u + \alpha'' u^2 + \alpha''' u^3.$$

The solutions of equation (3), which have just been found, are as follows:

1°. $K =$ arbitrary constant, $U = \alpha u^2 + \alpha' u + \alpha''$, $V = -\alpha v^2 + \alpha' v - \alpha''$.

2°. $K = \pm 2$, $U = \alpha u^2 + \alpha' u + \alpha''$, $V =$ arbitrary function of v.

3° (a). $K = \pm 1$, $U = \alpha u^3 + \alpha' u^2 + \alpha'' u + \alpha'''$, $V = \alpha v^3 - \alpha' v^2 + \alpha'' v - \alpha'''$.

3° (b). $K = \pm 1$, $U = \alpha u^2 + \alpha' u + \alpha'' + \dfrac{\alpha'''}{u}$, $V = -\alpha v^2 + \alpha' v - \alpha'' + \dfrac{\alpha'''}{v}$.

II. *Associates of the* BONNET *surfaces.*

§ 6. *Solution* 1° *of equation* (3). The surfaces arising from the first solution of equation (3), § 4, are the plane and cylindrical surfaces.*

The plane corresponds to the linear element.

$$ds^2 = (u+v)^K \left(\frac{du^2}{\alpha u^2 + \alpha' u + \alpha''} + \frac{dv^2}{-\alpha v^2 + \alpha' v - \alpha''} \right)$$

for any value of K, the form of the lines of reference depending of course upon the value of this constant.

*The proof of these statements follows from equations (2) and (3) and the values of the direction cosines of the normals to the surface found in the usual way.

For $K=2$, one set is composed of trochoids, and the other of logarithmic curves, which include the catenary as a particular case.* Cylindrical surfaces correspond to the linear element above when $K=0$.

§ 7. *Solution* 2° *of equation* (3). In discussing the surfaces arising from the solution of equation (3) when $K=-2$, BONNET found that two cases should be considered, depending upon whether the constant α which appears in the expression for U vanishes or not.

The expressions † for the linear elements and the cartesian coördinates of the surfaces in the two cases, as found by him, are respectively

$$(6)\qquad ds^2=\frac{1}{(u+v)^2}\left[\frac{du^2}{\alpha' u-u^2}+\frac{dv^2}{V}\right],$$

$$x=\frac{2\sqrt{\alpha' u-u^2}}{\alpha'(u+v)}, \qquad z=-\frac{2\sqrt{v^2+\alpha' v}}{\alpha'(u+v)}\sin V_3,$$

$$y=-\frac{2\sqrt{v^2+\alpha' v}}{\alpha'(u+v)}\cos V_3, \qquad V_3=\int\frac{4\sqrt{v^2+\alpha' v-V}}{2(v^2+\alpha' v)\sqrt{V}}dv;$$

$$(7)\qquad ds^2=\frac{1}{(u+v)^2}\left(\frac{du^2}{u}+\frac{dv^2}{V}\right),$$

$$x=2\sqrt{\frac{u}{\alpha'}}\frac{1}{(u+v)}, \qquad z=-2\sqrt{\frac{v}{\alpha'}}\frac{\sin V_3}{(u+v)},$$

$$y=-2\sqrt{\frac{v}{\alpha'}}\frac{\cos V_3}{(u+v)}, \qquad V_3=\int\frac{\sqrt{\alpha' v-V}}{2v\sqrt{V}}dv.$$

BONNET shows that the surfaces defined by (6) and (7) are characterized by having lines of curvature whose geodesic curvature is constant.‡ According to him, the surfaces defined by (6) are the envelopes of a variable sphere, whose center describes a plane curve, while its surface always passes through two fixed points, real or imaginary, which lie symmetrically with regard to the plane of the centers. In the case of the surfaces defined by (7), the two fixed points are coincident and lie in the plane of centers.

The cartesian coördinates (x_1, y_1, z_1) of the surfaces associated with (6) and (7) can be obtained by solving the system of equations, §

* See § 8.

† BONNET, loc. cit., p. 150. He gives only the expressions for the first case, but the others are easily derived from them.

‡ RIBAUCOUR has shown that the only surfaces having this property are cones, cylinders, and surfaces of revolution, together with surfaces obtained from these by inversion. (See DARBOUX, loc. cit., vol. 3, p. 122.) The surfaces (6) are the inverse of cones and surfaces of revolution, and (7) the inverse of cylindrical surfaces.

§ DARBOUX, loc. cit., vol. 2, p. 244.

$$\lambda\frac{\partial\theta_1}{\partial u}=\frac{\partial\theta}{\partial u},\qquad \lambda\frac{\partial\theta_1}{\partial v}=-\frac{\partial\theta}{\partial v},$$

where λ, which has its usual meaning, is obtained in the one case, from the expression for the linear element in (6), and in the other (7), and θ and θ_1 are to be replaced by x, x_1, y, y_1 and z, z_1 successively, the values of x, y and z being taken from (6) or (7), as the case may be.

We have, then, for the surfaces associated with (6),

$$ds_1^2=(u+v)^2\left[\frac{du^2}{\alpha' u+u^2}+\frac{dv^2}{V}\right],$$

and

$$x_1=\frac{2}{\alpha'}\sqrt{\alpha' u-u^2}\left(v+\frac{\alpha'}{2}\right)-\frac{\alpha'}{2}\operatorname{vers}^{-1}\frac{2u}{\alpha'},$$

$$(8)\qquad y_1=-\frac{2}{\alpha'}\sqrt{v^2+\alpha' v}\,(u+v)\cos V_3+\frac{4}{\alpha'}\int\sqrt{v^2+\alpha' v}\cos V_3\,dv,$$

$$z_1=\frac{2}{\alpha'}\sqrt{v^2+\alpha' v}\,(u+v)\sin V_3-\frac{4}{\alpha'}\int\sqrt{v^2+\alpha' v}\sin V_3\,dv,$$

where

$$V_3=\frac{\alpha'}{2}\int\frac{\sqrt{v^2+\alpha' v}-V}{(v^2+\alpha' v)\sqrt{V}}dv.$$

Likewise for those associated with (7), we have

$$ds_1^2=(u+v)^2\left(\frac{du^2}{\alpha' u}+\frac{dv^2}{V}\right),$$

$$x_1=2\sqrt{\frac{u}{\alpha'}}\left(\frac{u}{3}-v\right),$$

$$(9)\qquad y_1=-2\sqrt{\frac{v}{\alpha'}}\,(u+v)\cos V_3+\frac{4}{\sqrt{\alpha'}}\int\sqrt{v}\cos V_3\,dv,$$

$$z_1=2\sqrt{\frac{v}{\alpha'}}\,(u+v)\sin V_3-\frac{4}{\sqrt{\alpha'}}\int\sqrt{v}\sin V_3\,dv,$$

where

$$V_3=\int\frac{\sqrt{\alpha' v}-V}{2v\sqrt{V}}dv.$$

It follows from a theorem due to DARBOUX,* that any surface, S_1, defined by (8) or (9), can be so placed with regard to its associate surface, S, defined by (6) or (7), that at corresponding points the normals to the surfaces and the tangents to the lines of curvature will be respectively parallel.

* Loc. cit., vol. 2, p. 243.

The lines $v = \text{const.}$ on the BONNET surfaces are circles, therefore the corresponding lines on (8) and (9), namely, one set of the lines of curvature, must be plane.

The problem of determining isothermic surfaces having one set of the lines of curvature plane has been solved in its most general form by DARBOUX.* Of three possible exceptional or limiting cases,† which DARBOUX did not consider, one has been treated by ADAM,‡ another leads to the BONNET surfaces, and the third to their associates. This problem, then, is completely solved by the addition of the surfaces defined by (8) and (9).

§ 8. *Discussion of the associates of the Bonnet surfaces.* The principle of inversion applied to the surfaces (8) and (9) gives little information with regard to them, but we do know of course that the resulting surfaces will be isothermic and will have spherical lines for one set of the lines of curvature.

A discussion of equations (8) and (9) shows that the plane lines of curvature on the surfaces defined by them lie in a system of planes which envelope a cylindrical surface, the form of the latter depending entirely upon the form of the arbitrary function V. These lines are in fact the intersections of the planes and a system of cylindrical surfaces depending upon v as a parameter, whose elements are perpendicular to the elements of the cylinder enveloped by the planes.

For the surfaces defined by (8), the cross-section of these cylinders are either trochoids with their base lines parallel to the elements of the enveloped cylinder or curves whose equations are of the form

$$x = a\sqrt{2by + y^2} - \frac{\alpha_1}{2}\log\left(y + b + \sqrt{2by + y^2}\right),$$

according as the constant α' appearing in (8) is real or imaginary. The plane lines on the surfaces (9) are nodal cubic curves.

§ 9. *Surfaces having both sets of the lines of curvature plane.* The surfaces corresponding to the linear element

$$ds^2 = (u + v)^2\left(\frac{du^2}{\alpha u^2 + \alpha' u + \alpha''} + \frac{dv^2}{\beta v^2 + \beta' v + \beta''}\right),$$

when

* Excluding surfaces of revolution, DARBOUX reduced the general problem to the simultaneous solution of the equations

$$\frac{\partial h}{\partial u} = Ue^h + U_1 e^{-h}, \qquad \frac{\partial^2 h}{\partial u^2} + \frac{\partial^2 h}{\partial v^2} = 0,$$

where $e^h = \sqrt{\lambda}$ and U and U_1 are independent functions of u. He solved the problem when $U + U_1 + 0$, and ADAM considered the case $U = U_1$. The case $U_1 = 0$ leads to the BONNET surfaces, and $U = 0$ to their associates (DARBOUX, loc. cit., vol. 4, p. 222).

† See preceding note.

‡ Annales scientifiques de l'Ecole Normal supérieure, series 3, vol. 10 (1893), p. 319.

$$4(\alpha+\beta)(\alpha''+\beta'')-(\alpha'-\beta')^2=0,$$

have plane lines for both sets of the lines of curvature, since on the corresponding BONNET surfaces both sets are circles.* It is clear that surfaces obtained from these by inversion will be isothermic and that the lines of curvature will be spherical. The condition imposed upon the constants is satisfied when

1°. $\alpha=-1, \beta\neq 1, \alpha'\neq\beta', \beta''\neq 0,$
2°. $\alpha=-1, \beta=1, \alpha'=\beta', \beta''\neq 0,$
3°. $\alpha=-1, \beta\neq 1, \alpha'=\beta', \beta''=0,$
4°. $\alpha=0, \beta\neq 0, \alpha'\neq\beta', \beta''\neq 0,$
5°. $\alpha=0, \beta=0, \alpha'=\beta'.$

All other cases lead to imaginary surfaces.

In each of these cases, the integration indicated in (8) or (9) can be performed, the resulting functions being trigonometric or logarithmic. Case 1° is the general case. The lines $v=$ const. are the same as in the general case for (8), and the lines $u=$ const. are transcendental plane curves.

Case 2° leads to a minimal surface, which was first derived by DARBOUX † in connection with the general subject of minimal surfaces.

In case 3° the expression for the linear element becomes

$$ds^2=(u+v)^2\left(\frac{du^2}{\alpha' u-u^2}+\frac{dv^2}{\beta v^2+\alpha' v}\right),$$

and the cartesian coördinates are

$$(10)\qquad \begin{aligned} x&=\frac{2}{\alpha'}\sqrt{\alpha' u-u^2}\cdot v-\int\frac{u\,du}{\sqrt{\alpha' u-u^2}},\\ y&=\frac{2}{\alpha'}\sqrt{\beta v^2+\beta' v}\cdot u-\int\frac{v\,dv}{\sqrt{\beta v^2+\beta' v}},\\ z&=\frac{2}{\alpha'}\sqrt{1-\beta}\cdot uv. \end{aligned}$$

The two systems of planes in which the lines of curvature lie envelope cylindrical surfaces whose elements are perpendicular to one another.

(α) If, however, the constant β is greater than unity, the only real points on the surface are along the lines $u=0$ and $v=0$.

(β) If $\beta=1$ the surface is a plane and the lines of reference are like those on the DARBOUX minimal surface of the preceding case.

* BONNET, loc. cit., pp. 151–2.
† Loc. cit., vol. 1, p. 315.

(γ) If $\beta = 0$, the lines $u = \text{const.}$ are nodal cubic curves and the lines $v = \text{const.}$ are trochoids.

(δ) If $0 < \beta < 1$, the lines of curvature are similar to those on the DARBOUX minimal surface.

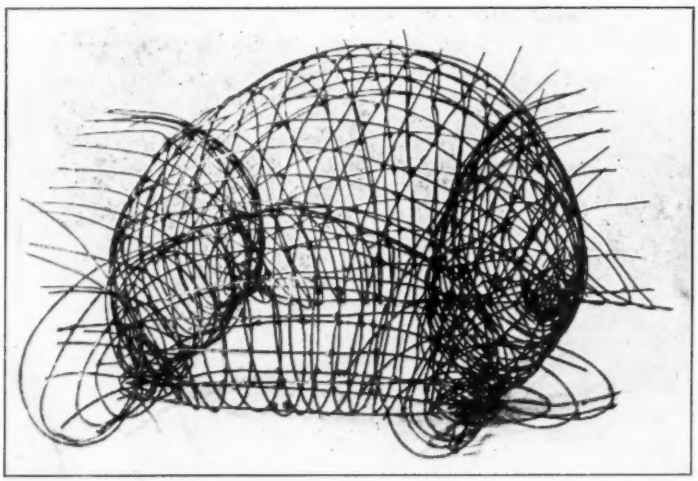
(ϵ) Finally, if β is negative, both sets of the lines of curvature are trochoids which have loops.

A model of one of these surfaces, the one corresponding to

$$ds^2 = (u+v)^2\left(\frac{du^2}{2u-u^2} + \frac{dv^2}{2v-v^2}\right),$$

has been made by the writer. This surface is divided into segments corresponding to the repeating parts of the lines of curvature. On the four sides of each segment to which others are joined are "rolls," formed by the loops of the lines of curvature, which taper uniformly from the middle of the side of the section, and reduce to points at the ends.

The whole surface is of such a form that it might be placed between two parallel planes, in such a way that one plane would touch the surface in a system



of isolated points (circular points on the surface, each being the crest of a segment) and so that the other would touch it along a double system of straight lines, which divide the plane into a net-work of squares, the sides of each being the locus of the points common to the plane and one segment of the surface.

In case 4° one set of the lines of curvature on the surface are nodal cubic curves, and the other consists of trochoids, or logarithmic curves. Finally, case 5° leads to the well known minimal surface of ENNEPER, an algebraic surface

of the ninth order, on which both sets of the lines of curvature are nodal cubic curves.

III. *Surfaces associated with the quadrics and with the sphere referred to confocal ellipses.*

§ 10. *Solution* $3°\,(a)$ *of equation* (3). BONNET * undertook to prove that the only surfaces having as their linear element the expression

$$ds^2 = (u+v)\left(\frac{du^2}{U} + \frac{dv^2}{V}\right),$$

when referred to lines of curvature, were the quadrics; but in solving (3), for the case $K=1$, he overlooked † solution $3°\,(a)$ for this value of K, and considered merely the solution $3°\,(b)$. Since however the solution overlooked leads to the sphere (as may be shown by computing the total curvature), his final results are not affected.

The lines of reference for the sphere are confocal spherical ellipses.‡

The expressions for the cartesian coördinates of the minimal § surface associated with the sphere may be obtained from its coördinates by a simple integration. They are as follows:

$$(11)\qquad \begin{aligned} x_1 &= \frac{1}{\sqrt{(a-b)(a-c)}} \log\left(\frac{\sqrt{a-v}+\sqrt{a-u}}{\sqrt{u-v}}\right),\\ y_1 &= \frac{1}{\sqrt{(b-a)(b-c)}} \log\left(\frac{\sqrt{b-v}+\sqrt{b-u}}{\sqrt{u-v}}\right),\\ z_1 &= \frac{1}{\sqrt{(c-a)(c-b)}} \log\left(\frac{\sqrt{v-c}+\sqrt{u-c}}{\sqrt{u-v}}\right), \end{aligned}$$

where $c<v<b<u<a$, and where the corresponding linear element is

$$ds_1^2 = \frac{1}{(u-v)}\left(\frac{du^2}{(u-a)(u-b)(u-c)} + \frac{dv^2}{-(v-a)(v-b)(v-c)}\right).$$

§ 11. *Solutions* $3°\,(b)$ *of equation* (3). BONNET has shown that the surfaces corresponding to this solution of (3) when $K=1$ are the quadrics.

We obtain the expressions for the cartesian coördinates of the associated surfaces by integration, from those obtained by him for the quadrics,‖ in the usual

* Loc. cit., p. 118–132.

† Loc. cit., p. 126.

‡ DARBOUX, loc. cit., vol. 1, p. 242.

§ DARBOUX, loc. cit., vol. 2, p. 244.

‖ Loc. cit., p. 115.

way. Corresponding to the linear element written in the form

$$ds_1^2 = \frac{1}{\mu^2 - \nu^2}\left(\frac{(\rho^2 - \mu^2)d\mu^2}{(\mu^2 - b^2)(c^2 - \mu^2)} + \frac{(\rho^2 - \nu^2)d\nu^2}{(b^2 - \nu^2)(c^2 - \nu^2)}\right),$$

we find the following systems of equations:

$$x_1 = \frac{\rho}{2bc}\log\left(\frac{\mu - \nu}{\mu + \nu}\right),$$

$$y_1 = \frac{1}{b}\sqrt{\frac{\rho^2 - b^2}{c^2 - b^2}}\sin^{-1}\sqrt{\frac{\mu^2 - b^2}{\mu^2 - \nu^2}},$$

$$z_1 = \frac{1}{2c}\sqrt{\frac{\rho^2 - c^2}{c^2 - b^2}}\log\left(\frac{\sqrt{c^2 - \nu^2} + \sqrt{c^2 - \mu^2}}{\sqrt{c^2 - \nu^2} - \sqrt{c^2 - \mu^2}}\right),$$

$$\nu^2 < b^2 < \mu^2 < c^2 < \rho^2;$$

$$x_1 = \frac{\rho}{2bc}\log\left(\frac{\mu - \nu}{\mu + \nu}\right),$$

$$y_1 = \frac{1}{b}\sqrt{\frac{\rho^2 - b^2}{c^2 - b^2}}\sin^{-1}\sqrt{\frac{\mu^2 - b^2}{b^2 - \nu^2}},$$

$$z_1 = \frac{1}{c}\sqrt{\frac{c^2 - \rho^2}{c^2 - b^2}}\sin^{-1}\sqrt{\frac{\mu^2 - c^2}{\mu^2 - \nu^2}},$$

$$\nu^2 < b^2 < \rho^2 < c^2 < \mu^2;$$

$$x_1 = \frac{\rho}{2bc}\log\left(\frac{\nu - \mu}{\nu + \mu}\right),$$

$$y_1 = \frac{1}{2b}\sqrt{\frac{\rho^2 - b^2}{c^2 - b^2}}\log\left(\frac{\sqrt{\nu^2 - b^2} + \sqrt{\mu^2 - b^2}}{\sqrt{\nu^2 - b^2} - \sqrt{\mu^2 - b^2}}\right),$$

$$z_1 = \frac{1}{c}\sqrt{\frac{c^2 - \rho^2}{c^2 - b^2}}\sin^{-1}\sqrt{\frac{c^2 - \mu^2}{\nu^2 - \mu^2}},$$

$$\rho^2 < b^2 < \mu^2 < c^2 < \nu^2.$$

The surfaces defined by the last three sets of equations correspond respectively to the ellipsoid, the hyperboloid of one sheet, and the hyperboloid of two sheets. The parameters can be eliminated in each of the three cases, giving an equation of the form $f(x_1, y_1, z_1) = 0$, where the non-algebraic functions involved are trigonometric or logarithmic.

PURDUE UNIVERSITY.